SPECIAL COLLECTOR'S EDITION

SCIENTIFIC AMERICAN

ALSO

DO WE LIVE IN AN INFINITE MULTIVERSE?

WONDERS OF THE COSMOS

Far-out tales from the edges of space and time

INSIDE

- Black hole at the beginning of time
- Searching for life on other worlds
- Why do stars explode?
- The largest structures in the universe

FALL 2017

© 2017 Scientific American

ADVERTISEMENT

THE ARCHIVES
Explore over 170 years of science history. Search any issue from 1845–present.
scientificamerican.com/all-access
SCIENTIFIC AMERICAN
The Evolution of EVOLUTION
How Darwin's Theory Survives, Thrives and Reshapes the World

Enjoy All-Access!

Read any issue, any year, on any device.

Receive 12 new issues (one year) of *Scientific American* in both print and digital. Plus, get full access to our award winning Archives, where you can explore any issue in our long history, 1845–present. Subscribe now. Go to:

scientificamerican.com/all-access

PRINT
12 issues filled with the latest advances in science.

SCIENTIFIC AMERICAN
A WEEKLY JOURNAL OF PRACTICAL INFORMATION, ART, SCIENCE, MECHANICS, CHEMISTRY, AND MANUFACTURES
NEW YORK, AUGUST 14, 1886.

DIGITAL
Online access to your subscription using any computer, tablet or smartphone.

Copyright © 2015 by Scientific American, a division of Nature America, Inc. All rights reserved.

SCIENTIFIC AMERICAN
ESTABLISHED 1845

Wonders of the Cosmos is published by the staff of SCIENTIFIC AMERICAN, with project management by:

EDITOR IN CHIEF AND SENIOR VICE PRESIDENT
Mariette DiChristina

DIGITAL CONTENT MANAGER **Curtis Brainard**
COLLECTIONS EDITOR
Andrea Gawrylewski

DESIGN DIRECTOR **Michael Mrak**
ISSUE DESIGNER **Lawrence R. Gendron**
PHOTOGRAPHY EDITOR **Monica Bradley**
ASSISTANT PHOTO EDITOR **Liz Tormes**
MANAGING PRODUCTION EDITOR **Richard Hunt**

COPY DIRECTOR **Maria-Christina Keller**
SENIOR COPY EDITOR **Daniel C. Schlenoff**
COPY EDITOR **Aaron Shattuck**
PREPRESS AND QUALITY MANAGER
Silvia De Santis

EDITORIAL ADMINISTRATOR **Ericka Skirpan**
SENIOR SECRETARY **Maya Harty**

SENIOR PRODUCTION MANAGER
Christina Hippeli
ADVERTISING PRODUCTION CONTROLLER
Carl Cherebin
PRODUCTION CONTROLLER
Madelyn Keyes-Milch

PRESIDENT
Dean Sanderson
EXECUTIVE VICE PRESIDENT
Michael Florek
EXECUTIVE VICE PRESIDENT, GLOBAL ADVERTISING AND SPONSORSHIP
Jack Laschever
PUBLISHER AND VICE PRESIDENT
Jeremy A. Abbate
ASSOCIATE VICE PRESIDENT, BUSINESS DEVELOPMENT
Diane McGarvey
HEAD, MARKETING AND PRODUCT MANAGEMENT
Richard Zinken
ASSOCIATE CONSUMER MARKETING DIRECTOR
Catherine Bussey
SENIOR CONSUMER MARKETING MANAGER
Lou Simone
MARKETING DIRECTOR, INSTITUTIONAL PARTNERSHIPS AND CUSTOMER DEVELOPMENT **Jessica Cole**
ONLINE MARKETING PRODUCT MANAGER
Zoya Lysak
MARKETING MANAGER **Marie Cummings**
MARKETING AND CUSTOMER SERVICE COORDINATOR
Christine Kaelin
HEAD, COMMUNICATIONS, USA
Rachel Scheer
DIRECTOR, INTEGRATED MEDIA **Jay Berfas**
SENIOR INTEGRATED SALES MANAGER
Matt Bondlow
DIRECTOR, GLOBAL MEDIA ALLIANCES
Ted Macauley
CUSTOM PUBLISHING EDITOR **Lisa Pallatroni**
RIGHTS AND PERMISSIONS MANAGER
Felicia Ruocco
SENIOR ADMINISTRATOR, EXECUTIVE SERVICES **May Jung**

COVER IMAGE BY RON MILLER; THIS PAGE NATIONAL GEOGRAPHIC CREATIVE/*Alamy*

ASTRONOMICAL radio telescopes at the Atacama Large Millimeter/submillimeter Array in Chile.

FROM THE EDITOR

Mysterious Universe

ON WHAT WOULD BECOME THE MOST WATCHED SHOW IN THE HISTORY OF PBS, astronomer Carl Sagan opined that "the Earth is a very small stage in a vast cosmic arena … our planet is a lonely speck in the great enveloping cosmic dark." The original *Cosmos* series, which aired nearly 40 years ago, instilled the haunting sense that there is much we don't understand about our universe. Indeed, the human mind may not even have the capacity to comprehend what gave rise to all matter since we are three-dimensional creatures living in a potentially 10-dimensional world.

The good news is that we are steadily gaining clarity. In this special edition, we present the most riveting discoveries about our cosmic home, from what we know of the origins of the universe and the cartography of space to the most bizarre phenomena and the enduring question: Are we alone in the universe?

Back to the problem of perception: Recent calculations by Niayesh Afshordi, Robert B. Mann and Razieh Pourhasan hint at a protouniverse—a fourth spatial dimension that existed even before the big bang and eventually exploded, casting a visual projection of that explosion. The resulting image could be the universe as we see it today (*page 4*). Another possibility is that during the initial expansion of the universe after the big bang, cosmic "bubbles" created multiple—if not an infinite number of—universes, of which ours is just one iteration (*page 30*).

Also in the mind-boggling department: Researchers speculate that entangled black holes could serve as wormholelike corridors around the universe (*page 46*). We are discovering freak molecules floating in space that could never exist on Earth (*page 60*) and enormous black holes that tear apart stars (*page 66*). Our universe certainly has a flare for the dramatic, though perhaps not when it comes to living things. Whether or not Earth is the only place that harbors animals, plants and microbes remains an open question (*page 74*), but we are scouring planets (*page 98*) and moons for signs of life (*page 78*).

The scale of these stories seems massive and shadowy at once, but Sagan might say that we are more capable and entitled to understand it than we know. "We are made of starstuff," he said. "We are a way for the cosmos to know itself." SA

Andrea Gawrylewski
Collections Editor
editors@sciam.com

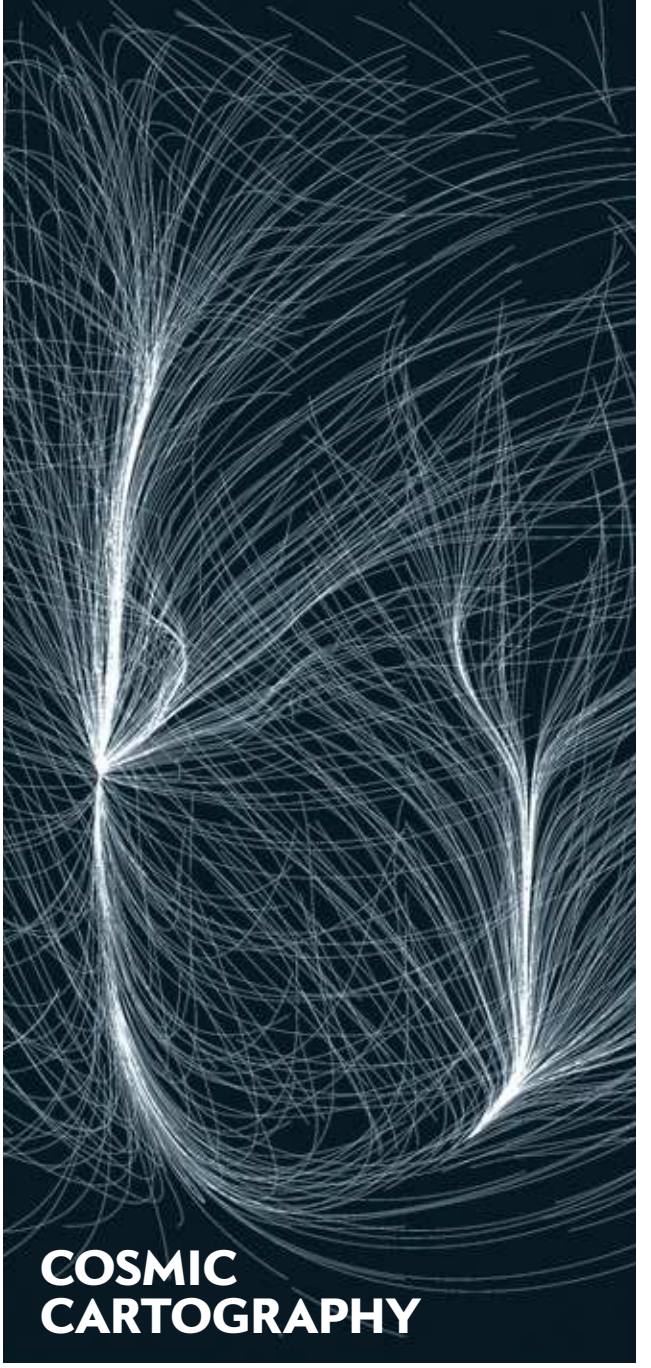
© 2017 Scientific American

SPECIAL EDITION

SCIENTIFIC AMERICAN

Volume 26, No. 4, Fall 2017

WONDERS OF THE COSMOS



4 The Black Hole at the Beginning of Time
Is the big bang, and all that came from it, a holographic mirage from another dimension? *By Niayesh Afshordi, Robert B. Mann and Razieh Pourhasan*

12 The First Starlight
Not long after the big bang's flash, all light left the cosmos. Astronomers are now solving the mystery of its return. *By Michael D. Lemonick*

20 The Puzzle of Dark Energy
Why is the expansion of the universe accelerating? After two decades of study, the answer is as mysterious as ever, but the questions have become clearer. *By Adam G. Riess and Mario Livio*

26 Ripples in Spacetime
A daring set of experiments has finally confirmed Einstein's 1916 prediction of gravitational waves. The discovery promises to open a new era in astrophysics. *By Clara Moskowitz*

30 The Quantum Multiverse
We may not live in a single universe but in a vast landscape of worlds known as the multiverse. The idea aligns with a 60-year-old interpretation of physics at the tiniest scales. *By Yasunori Nomura*

38 Our Place in the Cosmos
The Milky Way turns out to be part of a massive supercluster of galaxies that forms one of the largest known structures in the universe. This discovery is only the beginning of a new effort to map the cosmos. *By Noam I. Libeskind and R. Brent Tully*

46 Black Holes, Wormholes and the Secrets of Quantum Spacetime
The weird quantum phenomenon of entanglement could produce shortcuts between distant black holes. *By Juan Maldacena*

© 2017 Scientific American

WILD PHENOMENA

52 Stellar Fireworks
Every year thousands of exploding stars appear in a bizarre assortment of forms. Astronomers want to know what makes them go boom. *By Daniel Kasen*

60 Extreme Molecules in Space
Astrochemists are discovering many compounds in the cosmos that cannot exist on Earth. *By Clara Moskowitz*

66 How to Swallow a Sun
New techniques reveal how supermassive black holes shred entire stars. *By S. Bradley Cenko and Neil Gehrels*

LIFE OFF OF EARTH

74 Cosmic (In)significance
To know whether life exists beyond Earth, we must come to terms with our own significance in the universe. Are we uniquely special or merely mediocre? *By Caleb Scharf*

78 The Search for Life on Faraway Moons
Moons orbiting distant exoplanets may account for most of the habitable locales in the galaxy. If only we could find them. *By Lee Billings*

84 Strange News from Another Star
Alien technology is probably not causing Boyajian's star to dim mysteriously. But alternative explanations are hard to come by. *By Kimberly Cartier and Jason T. Wright*

90 Under the Sea of Enceladus
Evidence mounts that Saturn's icy moon harbors active hydrothermal vents, making it one of the hottest places to look for life beyond Earth. *By Frank Postberg, Gabriel Tobie and Thorsten Dambeck*

98 How to Search for Life on Mars
Experiments now under development could finally answer one of the most profound questions in science: Does life exist off of Earth? *By Christopher P. McKay and Victor Parro García*

104 Better Than Earth
Planets quite different from our own may be the best homes for life in the universe. *By René Heller*

WONDERS OF THE COSMOS

DEPARTMENTS

1 FROM THE EDITOR
Mysterious Universe

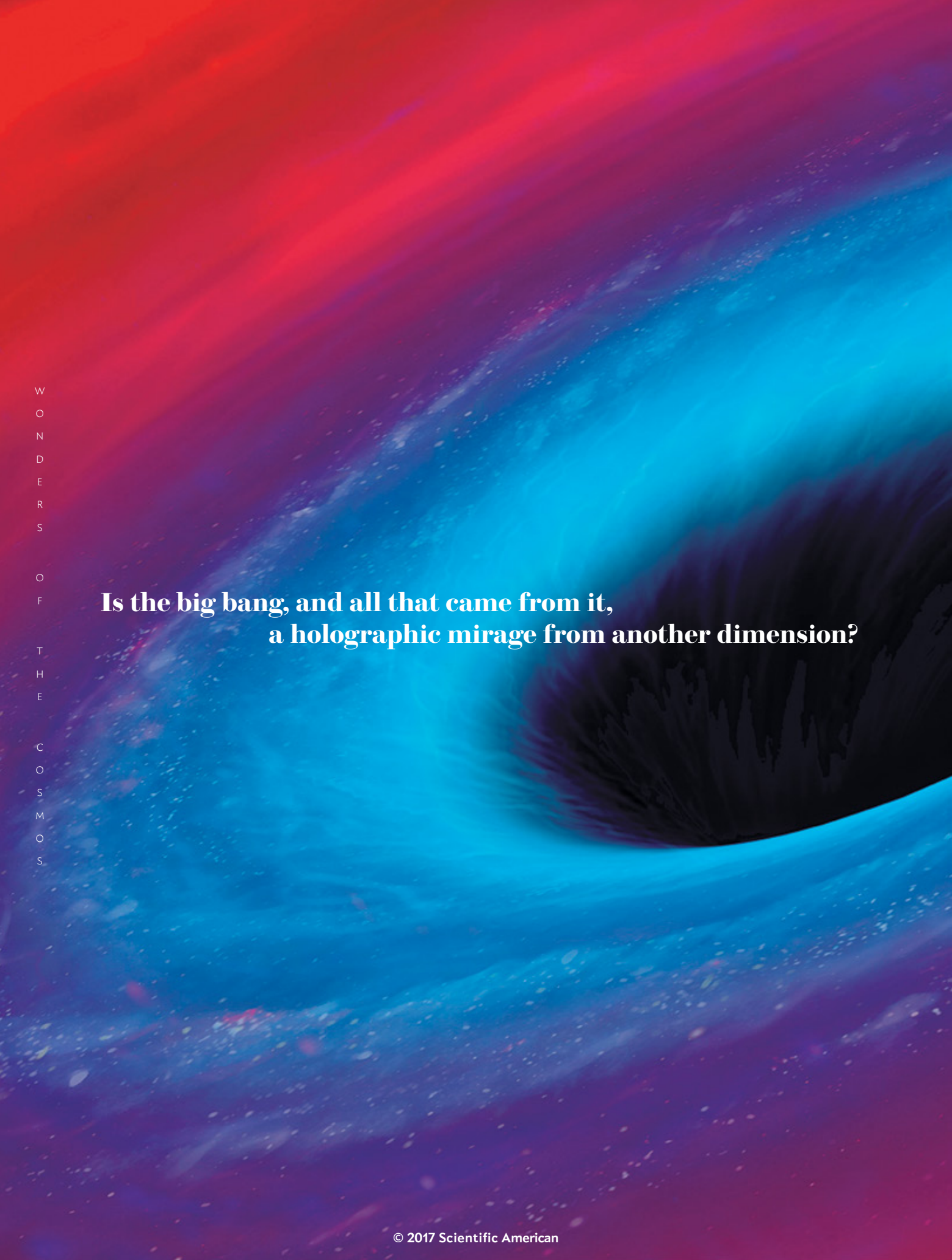
112 END NOTE
Long Live Hubble *By Katie Peek*

Articles in this special issue are updated or adapted from previous issues of *Scientific American* and from ScientificAmerican.com.

Copyright © 2017 Scientific American, a division of Nature America, Inc. All rights reserved.

Scientific American Special (ISSN 1936-1513), Volume 26, Number 4, Fall 2017, published by Scientific American, a division of Nature America, Inc., 1 New York Plaza, Suite 4500, New York, NY 10004-1562. Canadian BN No. 127387652RT; TVQ1218059275 TQ0001. To purchase additional quantities: U.S., $13.95 each; elsewhere, $17.95 each. Send payment to Scientific American Back Issues, 1 New York Plaza, Suite 4500, New York, NY 10004-1562. Inquiries: fax 212-355-0408 or telephone 212-451-8415. Printed in U.S.A.

© 2017 Scientific American



Is the big bang, and all that came from it, a holographic mirage from another dimension?

© 2017 Scientific American

THE BLACK HOLE AT THE BEGINNING OF TIME

By Niayesh Afshordi, Robert B. Mann and Razieh Pourhasan

© 2017 Scientific American

IN HIS ALLEGORY OF THE CAVE, GREEK PHILOSOPHER PLATO DESCRIBED PRISONERS who have spent their entire lives chained to the wall of a dark cavern. Behind the prisoners lies a flame, and between the flame and prisoners parade objects that cast shadows onto a wall in the prisoners' field of view. These two-dimensional shadows are the only things that the prisoners have ever seen—their only reality. Their shackles have prevented them from perceiving the true world, a realm with one additional dimension to the world that they know, a dimension rich with complexity and—unbeknownst to the prisoners—capable of explaining all that they see.

Plato was on to something.

We may all be living in a giant cosmic cave, created in the very first moments of existence. In the standard telling, the universe came into being during a big bang that started from an infinitely dense point. But according to recent calculations that we have carried out, we may be able to track the start of the universe back to an era before the big bang—an era with an additional dimension of space. This protouniverse may have left visible traces that upcoming astronomical observations could uncover.

The universe appears to us to exist in three dimensions of space and one of time—a geometry that we will refer to as the "three-dimensional universe." In our scenario, this three-dimensional universe is merely the shadow of a world that has *four* spatial dimensions. Specifically, our entire universe came into being during a stellar implosion in this suprauniverse, an implosion that created a three-dimensional shell around a four-dimensional black hole. Our universe is that shell.

Why would we postulate something that sounds, on the face of it, so absurd? We have two reasons. First, our ideas are not idle speculation—they are firmly grounded in the mathematics that describe space and time.

Over the past couple of decades physicists have developed a rich theory of holography, a set of mathematical tools that allows them to translate descriptions of events in one dimension to the physics of a different dimension. Researchers can solve relatively straightforward equations of fluid dynamics in two dimensions and use those solutions to understand what is going on in a much more complicated system—for example, the dynamics of a three-dimensional black hole. Mathematically, the two descriptions are interchangeable—the fluid serves as a perfect analogue for the extraordinary black hole.

The success of holography has convinced many scientists that more is at work here than a simple mathematical transformation. Perhaps the boundaries between dimensions are less stable than we thought. Perhaps the rules of the cosmos are written in another set of dimensions and translated into the three we perceive. Perhaps, like Plato's prisoners, our personal circumstances have tricked us into believing the world is three-dimensional when in fact a deeper understanding of what we perceive will come only when we look for explanations in the fourth dimension.

There is a second reason that our four-dimensional universe is worth thinking about. A close study of this universe could help us understand deep questions about

IN BRIEF

Cosmologists have detailed a remarkably accurate description of the history of the universe. But a few profound questions seem to defy all attempts at understanding.
One of these mysteries is the nature of the big bang itself—the sudden, violent origin of our universe from a point of infinite density.
The authors have developed ideas that would explain how the big bang came to be. They imagine that it emerged as a consequence of the formation of a black hole in a higher-dimensional universe. This theory provides answers to a number of difficult questions. It could also be tested.

PRECEDING PAGES: MARK GARLICK *Getty Images (artist's conception)*

© 2017 Scientific American

BIG BURST: A large stellar explosion can generate a black hole and a cloud of gas and dust called a supernova remnant (*red at left*). In a higher-dimensional universe, such an explosion could create our three-dimensional cosmos.

COURTESY OF NASA/JPL-CALTECH

the origin and nature of the cosmos. Consider, for example, the big bang, the primordial flash that brought our universe into existence. Modern cosmology holds that the big bang was immediately followed by "inflation"—a period of rapid expansion of space in which the early universe increased its volume by a factor of 10^{78} (or more). Yet this expansion provides no insight into what caused the big bang. Our four-dimensional universe, in contrast, gives us an answer to the ultimate mystery: Where did the universe come from?

THE KNOWN AND UNKNOWN COSMOS

OUR INVESTIGATIONS into the four-dimensional universe came about because of the problems that we have had contemplating the three-dimensional one. Modern cosmology has been fantastically successful, but its successes belie deep and complex mysteries that may lend themselves to a holographic explanation.

Cosmologists can describe the history of the entire universe—from the present day all the way back to a fraction of a fraction of a second after the big bang—using only a few equations (chief among them the ones provided by Albert Einstein) and five independent numbers, or parameters. These parameters include the densities of ordinary matter, dark matter and dark energy (more on these in a moment), along with the amplitude and shape of quantum fluctuations in the early universe. This model—the Lambda Cold Dark Matter (Λ-CDM) cosmological paradigm—describes hundreds (if not thousands) of observational data points, covering scales from a million light-years to 10 billion light-years across, right up to the edge of our observable universe.

But these observational successes do not mean our task is complete. The story of the universe is pocked with troublesome holes. We are confronted by fundamental questions about the nature of the cosmos—problems that we have not, as of yet, been able to answer.

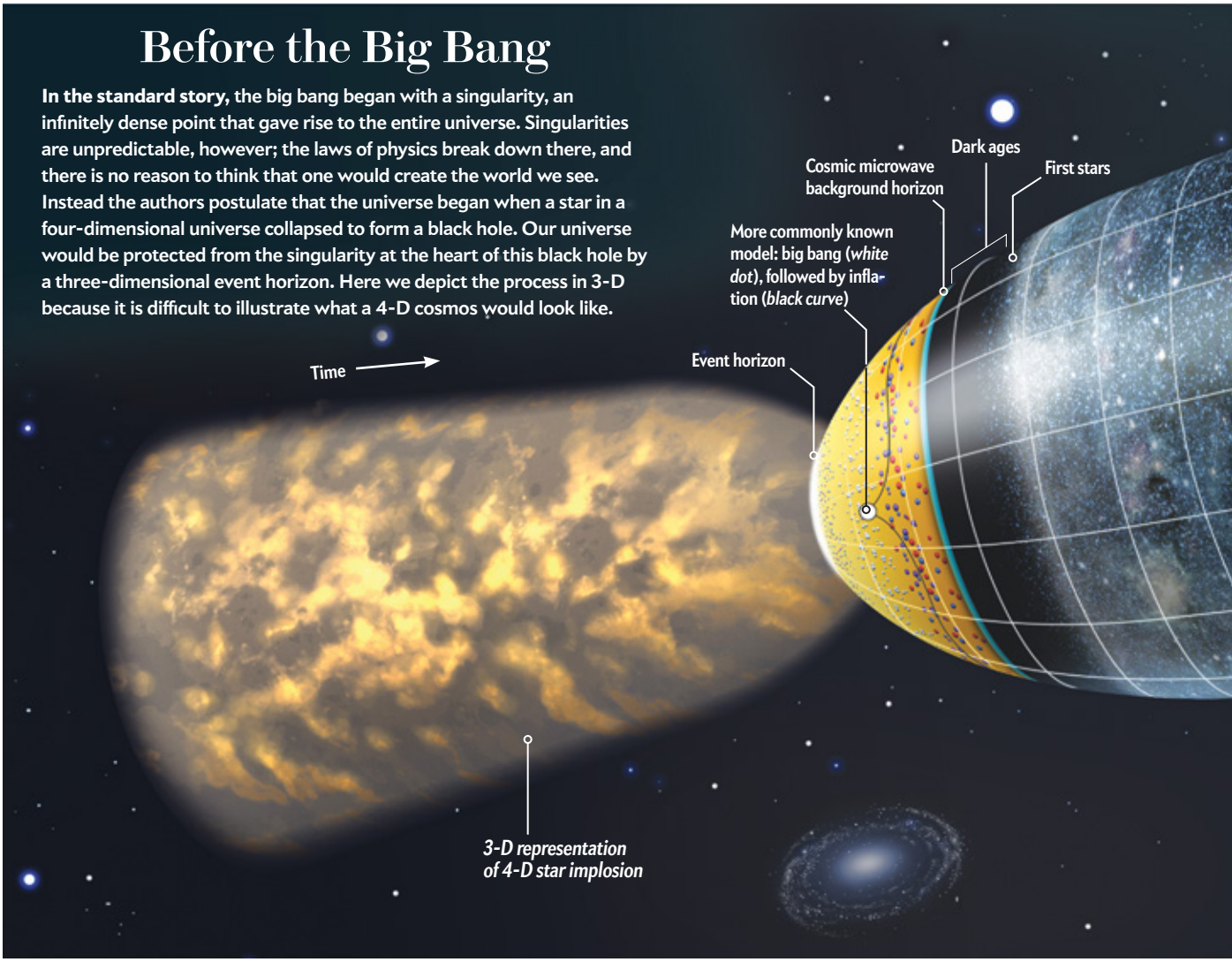
Problem 1: We don't understand the five parameters.

We do not have a satisfactory explanation for the origin of the five parameters of the Λ-CDM model, some of which must be very precisely chosen to agree with observations. Consider the density of matter and energy in the universe. Only a few decades ago astronomers believed that ordinary matter—the elements that make up the periodic table—would be the dominant form of mass-energy. Cosmological observations have radically revised this picture (and secured three Nobel Prizes along the way). We now know that the density of ordinary matter is only 5 percent of the universe's total energy density.

Another 25 percent comes in the form of dark matter, an unknown form of matter whose existence is inferred from its gravitational attraction. And 70 percent of the universe is made of dark energy, the mysterious stuff that is causing the expansion rate of our universe to speed up instead of slowing because of gravitational attraction. What are dark matter and dark energy, and why do they make up 25 and 70 percent of the universe, respectively? We do not know.

Perhaps answers would come if we better understood the big bang—the abrupt origin of space and time in a hot

© 2017 Scientific American


Before the Big Bang
In the standard story, the big bang began with a singularity, an infinitely dense point that gave rise to the entire universe. Singularities are unpredictable, however; the laws of physics break down there, and there is no reason to think that one would create the world we see. Instead the authors postulate that the universe began when a star in a four-dimensional universe collapsed to form a black hole. Our universe would be protected from the singularity at the heart of this black hole by a three-dimensional event horizon. Here we depict the process in 3-D because it is difficult to illustrate what a 4-D cosmos would look like.
Time
Dark ages
Cosmic microwave background horizon
First stars
More commonly known model: big bang (white dot), followed by inflation (black curve)
Event horizon
3-D representation of 4-D star implosion

plasma of radiation and particles at a temperature above 10^{27} degrees. It is very difficult to imagine how a situation like the universe in the moments after the big bang could lead to what we observe today—a cosmos of nearly uniform temperature and with a flat, large-scale spatial geometry (in which the angles of triangles sum up to 180 degrees).

Cosmic inflation might be the best idea we have for understanding the large-scale structure of the universe. Inflation would tend to "flatten" the universe, smoothing out any curved regions of spacetime, and bring it to a uniform temperature. Like a cosmic magnifier, inflation also amplifies tiny quantum fluctuations in energy density to cosmic size during this process. These fluctuations in turn become the seeds for the growth of structures such as galaxies, stars, planets and even living organisms such as ourselves.

Inflation is generally regarded as a very successful paradigm [*see box above*]. For decades cosmologists have been checking on inflation's predictions by observing the cosmic microwave background (CMB) radiation, a cosmic record of density fluctuations in the early universe. Observations by the Planck satellite confirm that our universe is flat (or very nearly so) and that it is uniform to better than one part in 60,000—as predicted by inflation. Furthermore, the observed amplitude and shape of primordial matter fluctuations are in broad agreement with how we would expect inflation to magnify the quantum vacuum.

Problem 2: We don't understand inflation completely.

We might ask what drove this inflation, which took a lot of energy. We imagine that, shortly after the big bang, the universe was filled with energy that takes the form of a hypothetical particle called the inflaton (pronounced "IN-flah-tahn"). The Higgs particle, discovered in 2012 by the Large Hadron Collider at CERN near Geneva, shares many properties with, and is a possible candidate for, the proposed inflaton. The inflaton could be responsible both for early accelerated expansion and for structure in our universe because the only significant density differences in the early universe are caused by the tiny quantum fluctuations in the inflaton field's energy.

Illustration by George Retseck

© 2017 Scientific American

But the inflaton does not solve our problems; it just pushes them back a step. The inflaton's properties, where it came from, and how to find it remain mysteries. We are not sure whether it really exists.

In addition, physicists do not understand how to naturally end inflation. If some kind of energy field drives an exponentially expanding universe, what would make that field suddenly turn off? And we lack a satisfactory description of the history of our cosmos before the inflationary era—those first trillionths of trillionths of trillionths of a second after the big bang.

Problem 3: We don't understand how it all began.

Cosmology's greatest challenge is understanding the big bang itself—the sudden, violent emergence of all space, time and matter from an infinitely dense point called a singularity. A singularity is an unimaginably bizarre thing, a point where space and time curve in on themselves, making it impossible to distinguish the future from the past. All the laws of physics break down. A singularity is a universe without order or rules. Out of a singularity could come anything that might logically exist. We have no reason to think that a singularity would generate a universe as ordered as the one we see.

We would expect the emergence of a universe from a singularity to be unthinkably chaotic, marked by huge temperature fluctuations from point to point. Furthermore, the magnifying power of inflation might be expected not to smooth everything out. In fact, if these fluctuations are too large, inflation may never get a chance to begin. The problems of a singularity cannot be solved by inflation alone.

Singularities are strange, but not unfamiliar. They also form at the centers of black holes, those collapsed remains of giant stars. All stars are nuclear furnaces that fuse lighter elements (primarily hydrogen) into heavier ones. This process of nuclear fusion powers a star for most of its life, but eventually the star exhausts all its nuclear fuel, and gravity takes over. A star at least 10 times more massive than our sun will collapse on itself before exploding as a supernova. If the star is even larger—15 to 20 solar masses or more—the supernova will leave behind a dense core that goes into a runaway collapse, contracting into a point of zero size—a black hole.

Black holes can be thought of as regions of space from which not even light can escape. Because the speed of light is the maximum speed attainable by any form of matter, the boundary of a black hole—a two-dimensional surface called the event horizon—is a point of no return: once stellar matter (or anything else) falls within this boundary, it is cut off from the rest of the universe and inexorably pulled toward the singularity at the center.

As with the big bang, the laws of physics break down at this singularity as well. Unlike the big bang, however, a black hole is surrounded by an event horizon. This surface acts like armored wrapping paper—it prevents any information about the singularity from leaking out. The event horizon of the black hole shields outside observers from the singularity's catastrophically unpredictable effects.

The event horizon effectively renders the singularity impotent, making it possible for the laws of physics to describe and predict all that we observe. Seen from a distance, a black hole would appear to be a simple, smooth and uniform structure, described only by its mass and angular momentum (and electric charge if it has any). Although physicists have recently raised some interesting questions about whether this conventional picture is consistent with quantum physics, the working assumption in cosmology is that black holes are cloaked by their event horizons.

In contrast, the big bang singularity (as commonly understood) is not cloaked. It has no event horizon. We would like

© 2017 Scientific American

TELESCOPES at the South Pole search for signals from the universe's first moments.

Whispers from Creation or Galactic Dust?

The quest to find gravitational waves from the big bang could turn up clues to how the universe began

Apart from the terrible weather hanging around from an unusually cold winter, the second week of March 2014 started like any other week. But then rumors started floating around in the cosmology community about an imminent announcement out of the Harvard-Smithsonian Center for Astrophysics. The rumors spread to Facebook, Twitter and the blogosphere by the weekend. Details began to emerge. This was not any ordinary announcement but rather the kind that, if correct, would happen once in a lifetime. It was something that most of us dreamed we could see only in a few decades if we were lucky, if at all.

The announcement from the BICEP-Keck collaboration was that an array of microwave telescopes located at the geographical South Pole had picked up startling patterns in the polarization of the cosmic microwave background. If these patterns were generated in the early universe, as the collaborators suggested was likely, they would confirm a 30-year-old prediction of the cosmic inflation theory: that the simplest models of inflation can generate an observable level of gravitational waves, comparable to density or temperature fluctuations in the early universe. It would also be our first direct evidence for the quantum nature of gravity, the most outstanding puzzle in theoretical physics over the past century.

Yet in science, as in life, things are rarely as simple as they first appear. For example, the simple inflationary models that predict observable levels of gravitational waves also suggest that hints of these waves should have been seen in the temperature fluctuations observed by the European Space Agency's Planck satellite. But they were not! Furthermore, microwave emission from dust in our galaxy tends to be polarized, which could confuse BICEP-Keck observations, at least to some extent.

As it turned out, further data released by the Planck team in late 2014 demonstrated that the original BICEP-Keck analysis had underestimated the degree of dust contamination in their maps. Dust in the Milky Way could in fact account for all of their signal. New experiments now planned or under way will improve significantly over the next decade, however. These should shed further light on the first moments after the big bang.

—N.A., R.B.M. and R.P.

to have a way to shield ourselves from the big bang's singularity and its catastrophic unpredictability, perhaps by something akin to an event horizon.

We have proposed one such scenario. It turns the big bang into a cosmic mirage. Our picture cloaks the singularity at the big bang just as an event horizon cloaks the singularity at the heart of a black hole. The cloaking protects us from the singularity's mercurial and nefarious effects.

EXTRADIMENSIONAL COLLAPSE

SUCH A CLOAK would differ from an ordinary event horizon in one critical way. Because we perceive that our universe has three spatial dimensions, the event horizon that cloaks the singularity at the heart of the big bang must also have three spatial dimensions—not just two. If we imagine that this event horizon also came about as a result of a cosmic collapse—just as a black hole's two-dimensional event horizon is formed by the

COURTESY OF STEFFEN RICHTER *Harvard University*

© 2017 Scientific American

collapse of a three-dimensional star—then the collapse would have to have taken place in a universe that had four spatial dimensions.

This kind of extradimensional scenario, in which the number of dimensions of space exceeds the obvious three, is an idea almost as old as general relativity itself. It was originally proposed by Theodor Kaluza in 1919 and expanded by Oskar Klein in the 1920s. Their idea was largely forgotten for more than half a century before being picked up by physicists studying string theory in the 1980s. More recently, scientists have used it to build a cosmology of so-called brane worlds.

The basic idea of a brane world is that our three-dimensional universe is a subuniverse embedded in a larger space of four or more spatial dimensions. The three-dimensional universe is called a brane, and the larger universe is called the bulk. All known forms of matter and energy are stuck to our three-dimensional brane like a movie projected on a screen (or the shadow reality for Plato's prisoners in the cave). The exception is gravity, which permeates all of the higher-dimensional bulk.

Let's think about the bulk supraniverse of four spatial dimensions that may have existed before the big bang. We can imagine that this bulk universe was filled with objects such as four-dimensional stars and four-dimensional galaxies. These higher-dimensional stars might run out of fuel, just as our three-dimensional stars do, and collapse into black holes.

What would a four-dimensional black hole look like? It would also have an event horizon, a surface of no return from which no light could escape. But instead of a two-dimensional surface, as we have in ordinary black holes, a four-dimensional black hole would generate an event horizon with three spatial dimensions.

Indeed, by modeling the collapsing death of a four-dimensional star, we observe, under a certain set of assumptions, that material ejected from the stellar collapse can form a slowly expanding three-brane surrounding this three-dimensional event horizon. Our universe is this three-brane—a hologram of sorts for a four-dimensional star collapsing into a black hole. The cosmic big bang singularity becomes hidden to us, locked away forever behind a three-dimensional event horizon.

IS THIS REAL?

OUR MODEL has a number of things going for it, starting with the fact that it eliminates the naked singularity that gave rise to the universe. But what of the other long-standing cosmological problems, such as the near flatness and high uniformity of the cosmos? Because the four-dimensional bulk universe could have existed for an infinitely long time in the past, any hot and cold spots in the bulk would have had plenty of time to come to equilibrium. The bulk universe would be smooth, and our three-brane universe would inherit this smoothness.

In addition, because the 4-D black hole would also appear to be nearly featureless, our emergent three-brane universe would likewise be smooth. The larger the mass of the four-dimensional star, the flatter the three-brane, and so the flatness of our universe is a consequence of its being residual detritus from the collapse of a heavy star.

In this way, our model of a holographic big bang resolves not only the main puzzles of uniformity and near flatness of standard cosmology without resorting to inflation but also nullifies the damaging effects of the initial singularity.

The idea may sound crazy, but there are several ways one might be able to test it. One way is by studying the cosmic microwave background radiation. Outside of our three-brane, we would expect there to be some extra four-dimensional bulk matter—something pulled close by the gravitational pull of the black hole. We can show that thermal fluctuations in this extra matter will create fluctuations on the three-brane that in turn distort the CMB by small but potentially measurable amounts. Our most recent calculations indicate that our model is broadly consistent with the newest data, but some simple inflationary models provide a better detailed fit.

Could our model have some missing pieces? For example, if the four-dimensional black hole is spinning (it is very common for black holes to spin), then our three-brane may not look the same in all directions. The large-scale structure of our universe would appear slightly different in different directions. Astronomers may also be able to find this directionality by studying subtle variations in the CMB sky.

Of course, even as the holographic big bang potentially resolves one giant question—the origin of our universe—it simultaneously raises a new set of mysteries. Foremost among them: Where did our universe's parent universe come from?

For an answer to this puzzle, we might again turn to Plato. When Plato's prisoners emerged from the cave, the light of the sun burned their eyes. It took them time to adjust to the brightness. At first, the prisoners were only able to make out shadows and reflections. Soon they could see the moon and the stars. Finally, they correctly concluded that the sun was "the author of all that we see"—day, night, season and shadow.

Plato's prisoners didn't understand the powers behind the sun, just as we don't understand the four-dimensional bulk universe. But at least they knew where to look for answers. SA

Niayesh Afshordi is an associate professor of physics and astronomy at the University of Waterloo and an associate faculty member in cosmology and gravitation at the Perimeter Institute for Theoretical Physics. Afshordi's research and teaching focus on astrophysics, cosmology and the physics of gravity.

Robert B. Mann is a professor and former chair of physics and astronomy at the University of Waterloo, an affiliate faculty member at the Perimeter Institute and a past president of the Canadian Association of Physicists. Mann's interests include black holes and quantum information.

Razieh Pourhasan is a postdoctoral fellow at the University of Iceland. She was born in Iran and received her Ph.D. at the Perimeter Institute and the University of Waterloo. Pourhasan focuses on applications of holography.

MORE TO EXPLORE

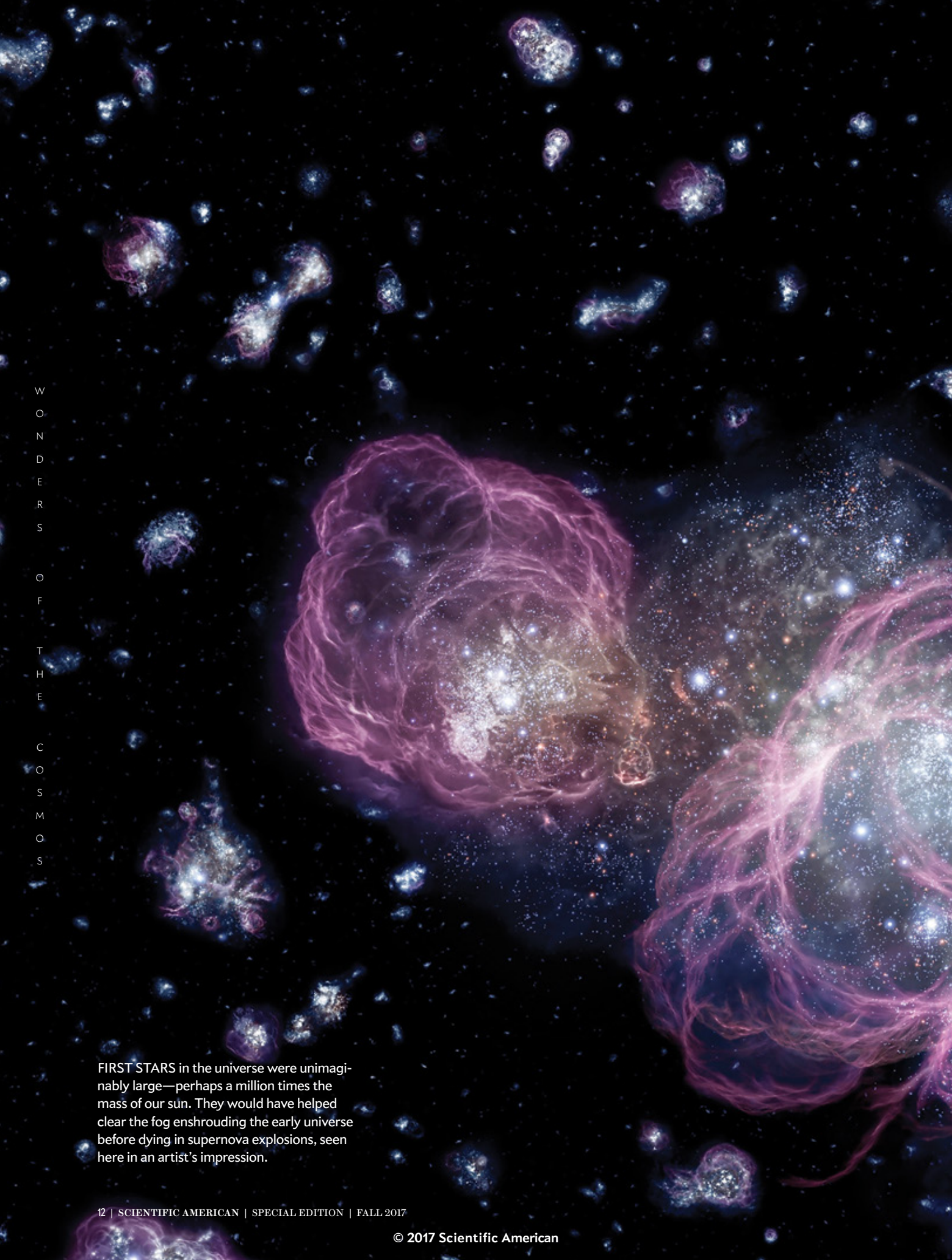
Out of the White Hole: A Holographic Origin for the Big Bang. Razieh Pourhasan, Niayesh Afshordi and Robert B. Mann in *Journal of Cosmology and Astroparticle Physics*, Vol. 2014, Article JCAP04(2014)005; April 2014. Available at http://dx.doi.org/10.1088/1475-7516/2014/04/005

A Holographic Big Bang? Niayesh Afshordi, Robert B. Mann and Razieh Pourhasan in *International Journal of Modern Physics D*, Vol. 24, No. 12, Article 1544029; 2015.

Cosmological Perturbations in the 5D Holographic Big Bang Model. Natacha Altamirano, Elizabeth Gould, Niayesh Afshordi and Robert B. Mann (in press). Preprint available at arXiv:1703.00954

scientificamerican.com/magazine/sa

© 2017 Scientific American



FIRST STARS in the universe were unimaginably large—perhaps a million times the mass of our sun. They would have helped clear the fog enshrouding the early universe before dying in supernova explosions, seen here in an artist's impression.

© 2017 Scientific American

THE FIRST STARLIGHT

Not long after the big bang's flash, all light left the cosmos. Astronomers are now solving the mystery of its return

By Michael D. Lemonick

© 2017 Scientific American

ABOUT 13.8 BILLION YEARS AGO, JUST 400,000 YEARS OR SO AFTER THE big bang, the universe abruptly went dark. Before that time, the entire visible universe was a hot, seething, roiling plasma—a dense cloud of protons, neutrons and electrons. If anyone had been there to see it, the universe would have looked like a pea soup fog, but blindingly bright. Around the 400,000-year mark, however, the expanding universe cooled enough for hydrogen atoms to form at last—an event known as recombination. The fog lifted, the universe continued to cool and everything quickly faded to black. After the unimaginable brilliance of the big bang and its immediate aftermath, the cosmos entered what astronomers call the dark ages of the universe.

And dark they were. For even when the first stars started to ignite, their light shone brightest in the ultraviolet portion of the spectrum—just the kind of light that the newly formed hydrogen gas tends to absorb. The universe traded its primordial hot, bright fog for one that was cool and dark.

Eventually this fog would lift, but how it did so is a question that has long baffled astronomers. Maybe it was accomplished by the first stars, whose intense light gradually but relentlessly broke the hydrogen apart in a process called reionization. Perhaps instead the energy for reionization came from the radiation that is generated by hot gas spiraling into giant black holes.

The key to figuring out how and when reionization took place is finding the oldest objects in the universe and trying to tease out their nature and their origins. When did the first stars turn on, and what were they like? How did individual stars assemble themselves into galaxies, and how did those galaxies form the supermassive black holes that lie at the core of nearly all of them? At what point in this progression from stars to galaxies to black holes did reionization take place? And was the process gradual or abrupt?

Astrophysicists have been asking many of these questions since the 1960s. Only recently, however, have telescopes and computer models gotten powerful enough to offer some answers. Computers have now simulated the emergence and evolution of the first stars in the universe. And telescopes are gathering telltale glimmers of light from less than half a billion years after the big bang—a time when the first galaxies were in their infancy.

SUPERSTARS

A DECADE OR SO AGO astronomers believed that they had a good handle on how the first generation of stars came to be. Immediately after recombination, the hydrogen atoms that filled the cosmos were spread uniformly through space. In contrast, dark matter, which physicists believe to be made of invisible particles that have not yet been identified, had already begun clumping together in clouds known as halos, averaging somewhere between 100,000 and one million solar masses. Gravity from these halos sucked in the hydrogen. As the gas became increasingly concentrated and heated up, it flared into light, creating the first stars in the universe.

In principle, this initial generation of giant stars, known to astronomers as Population III stars, could have broken up the gaseous veil of hydrogen atoms and reionized the universe. But much depends on the exact characteristics of these stars. If they were not bright enough or did not live long enough, they

IN BRIEF

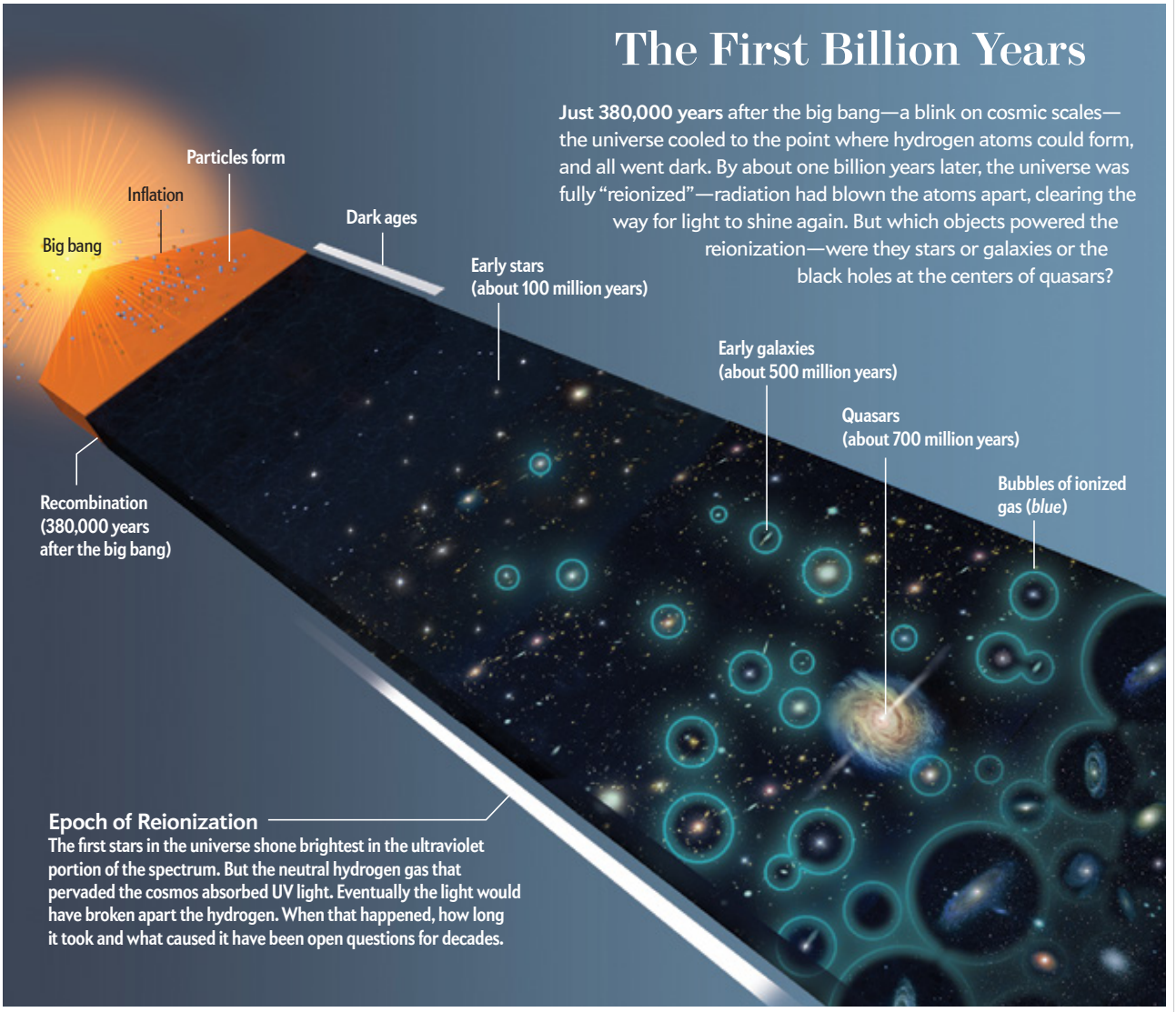
The universe's very first stars and galaxies were not like the objects we see today. Astronomers are reaching back in time to probe how the first objects in the universe came to be.

They are particularly interested in what caused the so-called reionization of the universe, when the neutral hydrogen atoms pervading the cosmos were broken up by light.

Observations and computer simulations suggest that the objects driving reionization could have been stars as massive as a million suns or the gaseous belches of enormous black holes.

PRECEDING PAGES: STOCKTREK IMAGES *Getty Images*

© 2017 Scientific American


The First Billion Years
Just 380,000 years after the big bang—a blink on cosmic scales—the universe cooled to the point where hydrogen atoms could form, and all went dark. By about one billion years later, the universe was fully "reionized"—radiation had blown the atoms apart, clearing the way for light to shine again. But which objects powered the reionization—were they stars or galaxies or the black holes at the centers of quasars?
Big bang
Inflation
Particles form
Dark ages
Early stars (about 100 million years)
Early galaxies (about 500 million years)
Quasars (about 700 million years)
Bubbles of ionized gas (blue)
Recombination (380,000 years after the big bang)
Epoch of Reionization
The first stars in the universe shone brightest in the ultraviolet portion of the spectrum. But the neutral hydrogen gas that pervaded the cosmos absorbed UV light. Eventually the light would have broken apart the hydrogen. When that happened, how long it took and what caused it have been open questions for decades.

would not be capable of finishing the job.

The characteristics of these stars depend strongly on their size. As of a decade ago, astronomers believed that they would be uniformly gigantic, each with roughly 100 times the mass of the sun [*see box on next page*]. The reason: As a clump of gas tries to collapse under gravity, it heats up. The heat creates so-called radiation pressure that opposes gravity; unless the star can shed some of this heat, the collapse will stall.

The first stars were made mostly of hydrogen, which is relatively terrible at shedding heat. (Stars like our sun also have small but critical traces of elements such as oxygen and carbon, which help them to cool.) As a result, a protostar in the early universe would continue to accumulate hydrogen gas, but the high pressure would prevent it from forming a dense core that would burst into a fusion reaction—one that drives much of the surrounding gas back out into space. The star would just gorge itself on more and more gas until it built a massive, diffuse core.

Now, however, says Thomas Greif, who created some of the most sophisticated simulations of early star formation while working as a postdoctoral fellow at Harvard University, "things look a bit more complicated." These newest simulations include not just gravity but also equations describing the feedbacks generated by increasingly pressurized hydrogen as the gas collapses. It turns out that the collapse of a hydrogen cloud can play out in many different ways. In some cases, the first stars could have been up to a million times as massive as the sun. In others, the collapsing cloud would have fragmented, creating several stars of just a few tens of solar masses.

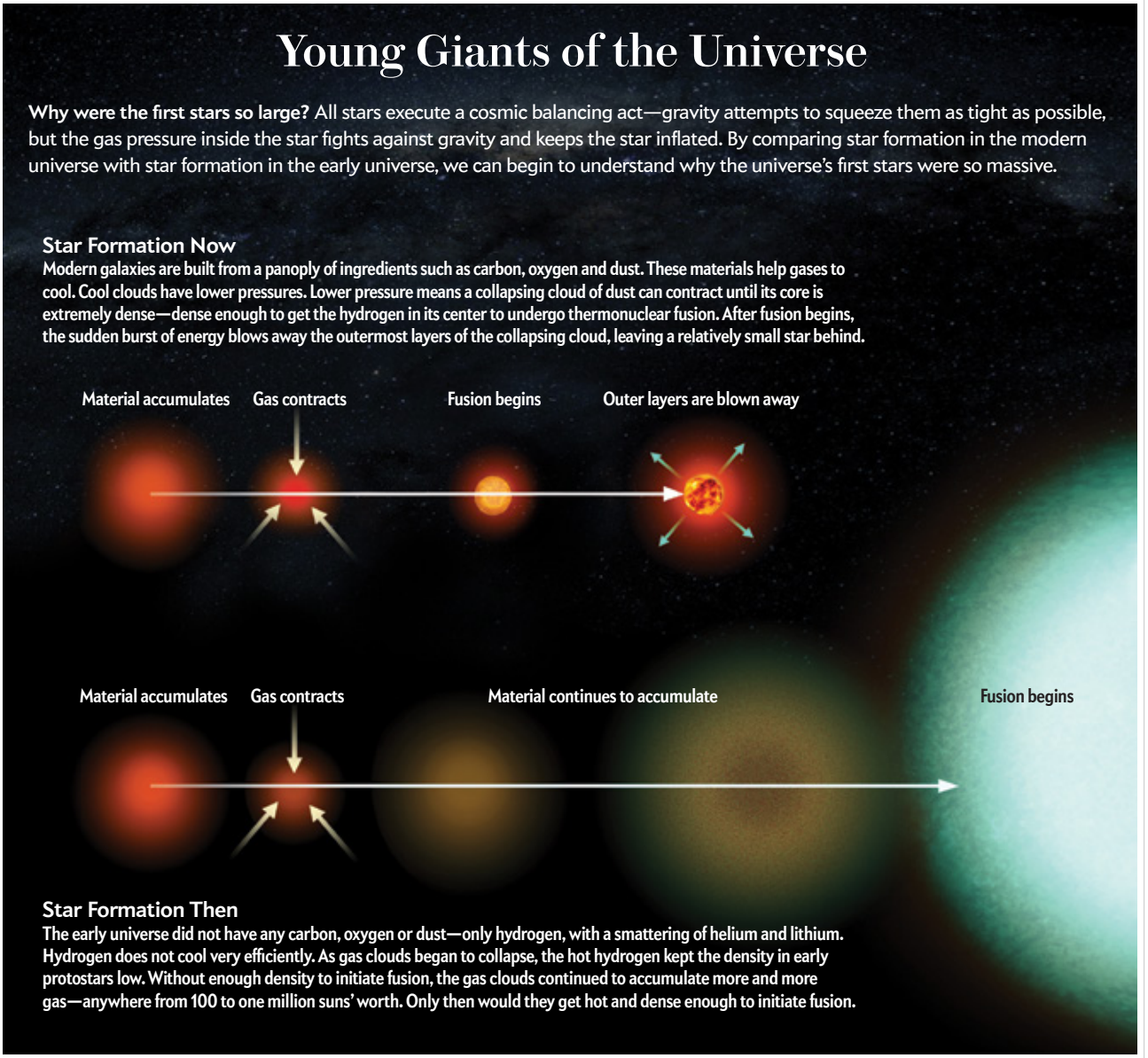
These enormous size differences imply correspondingly huge variations in the possible lifetimes of the first stars—and therefore in the onset and duration of the epoch of reionization. Giant stars of 100 solar masses or more are the rock 'n' rollers of astronomy: they live fast and die young. Smaller stars would churn through their nuclear fuel more slowly, implying that if stars were responsible for reionization, the process would have played out over many hundreds of millions of years [*see box above*].

Illustrations by Moonrunner Design

© 2017 Scientific American

Young Giants of the Universe

Why were the first stars so large? All stars execute a cosmic balancing act—gravity attempts to squeeze them as tight as possible, but the gas pressure inside the star fights against gravity and keeps the star inflated. By comparing star formation in the modern universe with star formation in the early universe, we can begin to understand why the universe's first stars were so massive.


Star Formation Now
Modern galaxies are built from a panoply of ingredients such as carbon, oxygen and dust. These materials help gases to cool. Cool clouds have lower pressures. Lower pressure means a collapsing cloud of dust can contract until its core is extremely dense—dense enough to get the hydrogen in its center to undergo thermonuclear fusion. After fusion begins, the sudden burst of energy blows away the outermost layers of the collapsing cloud, leaving a relatively small star behind.
Material accumulates
Gas contracts
Fusion begins
Outer layers are blown away
Material accumulates
Gas contracts
Material continues to accumulate
Fusion begins
Star Formation Then
The early universe did not have any carbon, oxygen or dust—only hydrogen, with a smattering of helium and lithium. Hydrogen does not cool very efficiently. As gas clouds began to collapse, the hot hydrogen kept the density in early protostars low. Without enough density to initiate fusion, the gas clouds continued to accumulate more and more gas—anywhere from 100 to one million suns' worth. Only then would they get hot and dense enough to initiate fusion.

BLACK LIGHTS

HOWEVER LARGE THEY WERE, all of these stars would have ended their existence in fiery supernovae before collapsing into black holes. And these black holes—perhaps more than the stars from which they came—may have then fueled the engine of reionization.

Black holes swallow nearby gas voraciously, and as the gas falls in, gravity compresses and heats it to temperatures of millions of degrees. The gas is so hot that while most of it eventually disappears into the black hole, some spews back out into space in the form of jets, which shine so brightly that the light can be seen halfway across the cosmos. We call these beacons quasars.

From the 1960s through the 1990s, quasars were really the only way to probe the early universe. At first, astronomers had no clue what they were. Quasars looked like nearby stars but had huge redshifts—a reddening of their light caused by the expansion of the universe. The impressive redshifts indicated that quasars were vastly farther away than any standalone star could possibly be and were thus vastly brighter as well. The first one ever found, 3C 273, had a redshift of 0.16, indicating that its light had begun traveling about two billion years ago.

"Then, very quickly," says Princeton University astrophysicist Michael A. Strauss, "people found quasars up to redshift 2"—a look-back time of more than 10 billion years. By 1991 Maarten Schmidt, James E. Gunn and Donald P. Schneider, working together at Palomar Observatory in California, had found a quasar with a redshift of 4.9, dating to 12.5 billion years before the present, or just a billion years and change after the big bang.

Yet analyses of the redshift 4.9 quasar found no evidence that its light was being absorbed by neutral hydrogen. Apparently the universe had already been reionized by the time the light from this

© 2017 Scientific American

quasar had begun its journey to Earth.

For most of the 1990s, no one was able to find a quasar any farther away than this one. The failure was not for a lack of powerful instruments—both the Hubble Space Telescope and the Keck telescopes at Mauna Kea in Hawaii came online in the early 1990s, significantly increasing astronomers' ability to see deep into the universe—but because quasars are rare to begin with. They erupt only from the most massive of supermassive black holes. And we can detect them only if their jets of gas happen to be aimed directly at us.

Moreover, those jets blast into existence only when a black hole is actively swallowing gas. For most black holes, that kind of activity peaked between a redshift of 2 and 3, when galaxies were more gas-rich, on average, than they are today. If you look farther out than that sweet spot in cosmic time, the number of quasars drops off rapidly.

It was not until 2000, when the Sloan Digital Sky Survey began methodically searching across an enormous swath of sky with the largest digital detectors ever built until that point, that the record was shattered in earnest. "The Sloan was just fabulously successful in finding distant quasars," says Richard Ellis, a professor of astrophysics at University College London. "They found something like 40 or 50 quasars beyond a redshift of 5.5." But the survey could not reach back much farther than a handful of quasars that it found between redshift 6 and 6.4, and even at that distance there was no sign of neutral hydrogen.

It was only with the discovery of a quasar at redshift 7.085, by the UKIRT Infrared Deep Sky Survey at Mauna Kea, that astronomers found small but significant amounts of ultraviolet-absorbing hydrogen obscuring the object's light. This quasar, known as ULAS J1120+0641, was shining about 770 million years after the big bang. It finally let astrophysicists dip a toe into the era of cosmic reionization—but just a toe because even this close to the big bang, most of the neutral hydrogen had already been destroyed.

Or maybe not. It is possible that this quasar sits in an unusually sparse region of leftover neutral hydrogen and that most other quasars at this distance would have been more completely shrouded. It is equally possible that ULAS J1120+0641 is in an especially dense region; maybe reionization was already complete pretty much everywhere else. Without more examples, astronomers cannot be sure, and the prospects of finding enough quasars at this distance to do a robust statistical study are slim.

ULAS J1120+0641 has plenty to tell astronomers anyway. For one thing, Ellis says, "the number of quasars is falling so steeply with distance that it's inconceivable that massive black holes are a major source of radiation that reionizes the universe." For another, the black hole that powers this particular quasar must have a billion suns' worth of mass to generate the amount of energy that makes it visible from so far away. "It's almost impossible to understand how it could have formed in the limited time that the universe had up to that point," Ellis says.

IT IS ALMOST IMPOSSIBLE TO UNDERSTAND HOW THE BLACK HOLE COULD HAVE GROWN SO LARGE IN THE LIMITED TIME THAT THE UNIVERSE HAD UP TO THAT POINT. YET FORM IT DID.

Nevertheless, form it did. Abraham Loeb, chair of Harvard's astronomy department, points out that if a first-generation star of 100 solar masses collapsed into a black hole a couple of hundred million years after the big bang, it could conceivably have grown to quasar proportions in the available time if conditions were just right. "But you would need to feed the black hole continuously," he says, and it is hard to imagine how you could do that. "They shine so brightly, they produce so much energy, that they expel the gas out of their vicinity." Without a nearby supply of gas, the quasar goes temporarily dark, allowing gas to accumulate again until it can flare back into life—and blow away its fuel supply once more. As this cyclical process repeats, Loeb says, "the black hole is able to grow only for a fraction of the time."

Yet black holes can also get bigger by merging with one another, and merging should accelerate their growth. In addition, Loeb and his co-authors suggested in a 2003 paper that those first black holes may have formed from stars that were not 100 solar masses, but one million solar masses. "This has become a popular idea," Loeb says, buttressed by the recent work on star sizes and by simulations such as Greif's. "And because these stars would shine as brightly as the entire Milky Way, you could, in principle, see them with the James Webb Space Telescope," the massive successor to the Hubble telescope that is currently scheduled to launch in 2018.

GALAXY QUEST

EVEN AS THE HUNT for distant quasars has more or less petered out, the search for galaxies closer and closer to the big bang has taken off. Perhaps the most important triggering event was an image called the Hubble Deep Field. It was made in 1995, when Robert Williams, then director of the Space Telescope Science Institute, used a perk of the office known as "director's discretionary time" to aim Hubble at an evidently blank patch of sky and let it stare for a cumulative 30 hours or so to pick up whatever faint objects might be there. "Some very serious astronomers told him it was a waste of time," recalls former director Matt Mountain, "that he wouldn't see anything."

In fact, the telescope picked up several thousand small, faint galaxies, many of them among the most distant ever seen. Later Deep Field images—made with Hubble's upgraded, infrared-sensitive Wide Field Camera 3, which is about 35 times more effective than its predecessor was—have found even more.

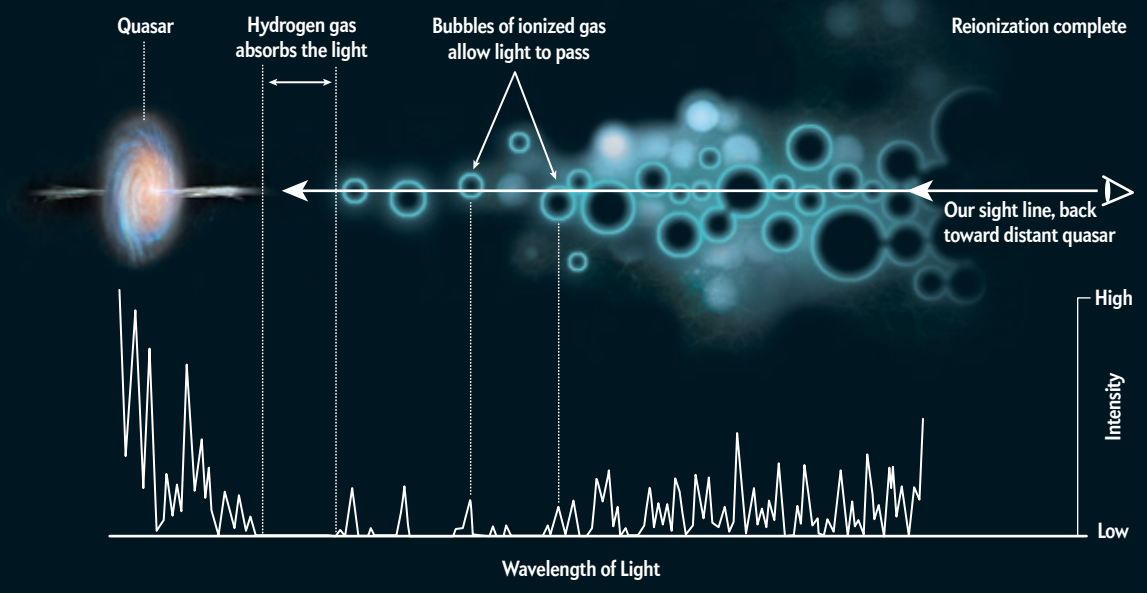
"We've gone from four or five galaxies with a redshift of 7 or more to more than 100," Daniel Stark, a professor of astronomy at the University of Arizona. One of these ultradistant galaxies, described by Stark, Ellis and several co-authors in a 2012 paper, appears to be at a redshift of no less than 11.9, which dates it to fewer than 400 million years after the big bang.

Like the record-holding quasar, these young galaxies can tell astronomers plenty about the distribution of intergalactic hydrogen at the time. When observers look at their output of ultraviolet light, a significant fraction of what they would expect to see is missing, absorbed by neutral hydro-

© 2017 Scientific American

Quasar Quest

Quasars are among the brightest objects in the early universe, beacons that astronomers can spot from more than 10 billion light-years away. As light from a quasar travels through the universe toward our telescopes, two things happen: First, its light gets stretched along the way by the expansion of the universe. In addition, intervening atomic hydrogen gas absorbs some of the light. Astronomers plot the absorption of light by wavelength to see how the prevalence of hydrogen gas has changed over time. They have found that isolated bubbles of ionized gas grew larger and more numerous as the universe evolved.



gen that surrounds them. That fraction drops gradually as they look at galaxies that are farther from the big bang—until, at about a billion years after the universe was born, the cosmos is fully transparent.

In short, not only did galaxies exist to provide a source for the ionizing radiation, they also reveal how the universe made the transition from neutral to fully ionized. Astronomical detectives have a smoking gun, and they have a victim. There is a catch, however. When they take the 100-odd galaxies found so far above a redshift of 7 and extrapolate across the entire sky, they do not come up with enough total ultraviolet radiation to ionize all the neutral hydrogen. The gun does not seem to have been powerful enough to do the job. The required energy could not have come from black holes, either; there just was not enough time for that many supermassive black holes to form.

Yet the answer may be relatively straightforward. Faint as they seem to us, the galaxies we are able to see at the very edge of Hubble's vision are presumably the brightest of their epoch. There must be many more galaxies at that distance that are simply too dim to see with any existing telescope. The assumption is reasonable enough that "I think most people now believe that galaxies do most of the work in reionizing the universe," Ellis says.

THE EINSTEIN CARD

AS FOR WHAT TRULY newborn galaxies look like and when they first turned on, "we're not there yet," Stark admits. "The galaxies we do see are fairly small, and they look much younger than galaxies that have been studied in detail [from] one billion to two billion years later." But these small, young galaxies already have as many as 100 million stars. And the mix of their colors (after you correct for the fact that their light is redshifted) suggests their stars are on average redder than one would expect to see in an infant galaxy.

"These objects," Stark says, "look like they've been forming stars for at least 100 million years already. Hubble has taken us close to the precipice, to where we'll see the first generation of stars. But it will take the James Webb Space Telescope to get us there."

Hubble has not exhausted its options, however. The telescope itself can see only to a certain faintness limit without taking absurdly long exposures. Yet the universe has supplied its own natural lenses that can boost Hubble's power. These so-called gravitational lenses take advantage of the fact that massive objects—in this case, clusters of galaxies—warp the space around them, distorting and sometimes magnifying the objects that lie far beyond.

In particular, says observer Marc Postman of the Space Telescope Science Institute, "we get a big amplification of any very distant galaxies that lie behind those clusters. They can be 10 or 20 times brighter than comparable unlensed galaxies." Post-

SOURCE: GEORGE DJORGOVSKI *California Institute of Technology* AND ABRAHAM LOEB *Harvard University*

© 2017 Scientific American

man is principal investigator for the Cluster Lensing and Supernova Survey with Hubble, a program that has used the technique to identify some 250 additional galaxies between redshift 6 and 8 and a handful more that may go up to redshift 11. From what they have seen so far, the results are consistent with those coming out of the various Deep Field surveys.

Recently Hubble went deeper still in a project called Frontier Fields, which started in 2013. Through the end of 2016, Mountain and other astronomers discovered images of distant galaxies 100 times fainter than anything previously seen. The galaxies lie behind six especially massive and powerful clusters. Each cluster was observed for 140 orbits of Hubble, totaling more than 100 hours. "That will let us probe deeper into the universe than anything we've ever seen," says Jennifer Lotz, the project's lead observer.

BURST SEARCH

YET ANOTHER KIND of cosmic beacon, meanwhile, could ultimately prove to be an even better probe of the early universe. When first discovered in the 1960s, gamma-ray bursts—short blasts of high-frequency radiation that pop up in random directions—were an utter mystery. Nowadays astronomers believe that many of them come from the deaths of very massive stars. As the stars collapse to form black holes, they spew jets of gamma rays out into space. When the jets slam into the surrounding clouds of gas, they trigger a secondary, bright afterglow of visible and infrared light that can be seen by conventional telescopes.

Here is how the observations work: When the orbiting Swift Gamma-Ray Burst observatory detects a gamma-ray flash, it swivels to point its onboard telescopes at the spot. At the same time, it beams the location's coordinates to ground-based observers. If their telescopes get there before the flash fades, astronomers can measure the afterglow's redshift and thus the redshift—and age—of the galaxy where the burst went off.

What makes the technique so valuable is that gamma-ray bursts make other cosmic objects look positively feeble. "For the first few hours," says Edo Berger, a Harvard astrophysicist who specializes in the bursts, "they probably outshine galaxies by a factor of a million, and they're 10 to 100 times brighter than quasars." You do not need a long exposure with Hubble to see them. In 2009 a telescope on Mauna Kea reliably measured the redshift of one burst at 8.2, putting it at 600 million years after the big bang.

The flash was so bright, Berger says, that it could have been seen out to a redshift of 15 or even 20, which would be less than 200 million years after the big bang, close to the time the very first stars may have been shining. And it is entirely plausible, he says, that those very massive stars would be exactly the kind to produce gamma-ray bursts as they die. In fact, Berger says, there is reason to think these first-generation stars would create such energetic gamma-ray bursts that they would appear brighter than the ones discovered so far, even though they would be farther away.

Unlike quasars, moreover, which occur only in galaxies with supermassive black holes, and unlike the galaxies that Hubble can see, which are the brightest tips of a grand galactic iceberg, gamma-ray bursts are just as powerful in tiny galaxies as they are in big ones. They provide, in other words, a much more representative sample of the universe at any given time.

The downside, Berger says: 99 percent of gamma-ray bursts are pointed away from Earth, and of the remaining one per day or so that are detected by satellites, only a minuscule fraction are at a high redshift. Gathering a representative sample of extremely high redshift bursts would therefore take a decade or more, and Swift probably will not last that long, Berger says. Ideally, he notes, someone should launch a successor satellite that could feed burst coordinates to the James Webb telescope or to the three 30-meter-class ground-based instruments that are expected to be operating within the next decade. Proposals to do so have so far failed to get the go-ahead from either NASA or the European Space Agency.

In any case, once the James Webb telescope and the next generation of gigantic ground-based telescopes begin observations, quasar hunters, galaxy surveyors and those who search for the telltale afterglows of gamma-ray bursts in other electromagnetic wavelengths will be able to catalogue much older and fainter objects than they can today. Their work will help to nail down exactly what was going on in the very early universe.

Radio astronomers, meanwhile, will be looking to a number of instruments—including the Murchison Widefield Array in Australia, the Precision Array for Probing the Epoch of Reionization in South Africa and West Virginia, the Square Kilometer Array to be built in Australia and South Africa, and the Low Frequency Array antennas located in several European countries—to map slowly disappearing clouds of neutral hydrogen during the first billion years of cosmic history.

The hydrogen itself emits radio waves. So in principle astronomers can look at those emissions at different epochs, each redshifted by a different amount, depending on how far away they are. They can then piece together snapshots of the hydrogen as it is gradually eaten away by high-energy radiation. Astronomers have used the Atacama Large Millimeter/submillimeter Array in the Chilean desert, for example, to detect carbon monoxide drifting in the space between the second-generation stars in immature galaxies a billion years after the big bang.

When cosmologists first detected the leftover electromagnetic radiation from the big bang in 1965, it galvanized them to try to understand the life history of the universe from its birth right through to the present. They are not quite there yet. But there is every reason to believe that by 2025, the 60th anniversary of that discovery, the last remaining blanks will finally be filled in. SA

Michael D. Lemonick is chief opinion editor for *Scientific American* and author, most recently, of *The Perpetual Now: A Story of Amnesia, Memory and Love*. For 21 years he was a science writer for *Time* magazine.

MORE TO EXPLORE

The Dark Age of the Universe. Jordi Miralda-Escudé in *Science*, Vol. 300, pages 1904–1909; June 20, 2003. Preprint available at http://arxiv.org/abs/astro-ph/0307396

The First Stars. Volker Bromm and Richard B. Larson in *Annual Review of Astronomy and Astrophysics*, Vol. 42, pages 79–118; May 19, 2004. Preprint available at http://arxiv.org/abs/astro-ph/0311019

The First Stars in the Universe and Cosmic Reionization. Rennan Barkana in *Science*, Vol. 313, pages 931–934; August 18, 2006. Preprint available at http://arxiv.org/abs/astro-ph/0608450

James Webb Space Telescope: http://jwst.nasa.gov/science.html

scientificamerican.com/magazine/sa

© 2017 Scientific American

The Puzzle of Dark Energy

Why is the expansion of the universe accelerating? After two decades of study, the answer is as mysterious as ever, but the questions have become clearer

By Adam G. Riess and Mario Livio

THE UNIVERSE IS GETTING BIGGER EVERY SECOND. GALAXIES ARE FLYING apart from one another, clusters of galaxies are zooming away from other clusters, and the empty space between everything is growing wider and wider. This much was known since the 1920s, when observations by Edwin Hubble and others revealed that the cosmos is expanding. But more recently, astronomers found that the process is speeding up—the pace of the expansion of the universe is rising, so that galaxies are receding from one another faster now than they were a moment ago. This is the startling realization that one of us (Riess), along with collaborators he co-led with Brian Schmidt of the Australian National University, came to in 1998 through measurements of distant supernovae explosions.

Illustration by Kenn Brown, Mondolithic Studios

© 2017 Scientific American

LIKE A RUBIK'S CUBE, the quandary of dark energy—the mystery of why the expansion of the universe is speeding up—is turning out to be hard to crack.

© 2017 Scientific American

The discovery agreed with findings from another team led by Saul Perlmutter of the University of California, Berkeley, using a similar method published the same year. The conclusion was unavoidable—something was causing the expansion of the universe to pick up speed. But what?

We give the name "dark energy" to whatever is causing the repulsive force that is apparently pulling the universe apart. After studying the situation for nearly two decades, the physical nature of dark energy remains almost as elusive today as it was 19 years ago. In fact, the latest observations only seem to further complicate the picture by showing hints of disagreement with the leading theories.

We are left with several pressing mysteries: What is dark energy? Why does it seem so much weaker than the most straightforward theories predict (and yet strong enough that we detect it)? What does the nature of dark energy mean for the future of the universe? And finally, do the strange characteristics of dark energy imply that our universe got its properties randomly—that in fact, our universe is just one of a vast multiverse containing untold versions of the cosmos, each with different traits and different strengths of dark energy?

An all-out assault to identify the nature of dark energy is under way, and prospects are bright for several new observatory projects to make progress soon. Within the next decade we hope to begin to answer these questions and comprehend the nature of cosmic acceleration—or resign ourselves to leaving some mysteries unsolved indefinitely.

WHAT IS DARK ENERGY?

SCIENTISTS HAVE A NUMBER of hypotheses for what might be driving the acceleration of the universe. The leading candidate arises from the nature of empty space. In quantum physics a vacuum is not "nothing"—rather it is teeming with pairs of "virtual" particles and antiparticles that spontaneously appear and annihilate one another within a tiny fraction of a second. As strange as it may sound, this sea of ephemeral particle pairs carries energy, and energy, just like mass, can produce gravity. Unlike mass, however, energy can create either an attractive or a repulsive gravity, depending on whether its pressure is positive or negative. The vacuum energy in empty space, according to theory, should have a negative pressure and thus may be the source of the repulsive gravity driving the accelerated expansion of the universe.

This idea is equivalent to the "cosmological constant," a term Albert Einstein added to his general relativity equations that represents a constant energy density throughout space. As the name implies, this hypothesis holds that the density of dark energy is constant—that is, unvarying—over space and time. So far the astrophysical evidence we have best fits with the cosmological constant explanation, with some discrepancies.

Alternatively, dark energy may be an energy field dubbed "quintessence" that pervades the universe, imbuing every point in space with a property that counteracts the pull of gravity. Physicists are familiar with fields—the everyday forces of electromagnetism and gravity act via fields (although these usually arise from localized sources and do not pervade all of space).

If dark energy is a field, it would not be a constant and so might change over time. In that case, dark energy might once have been stronger or weaker than it is now and could have affected the universe differently at different times. Likewise, its strength and impact on the evolution of the universe might alter in the future. In the so-called freezing-field version of this idea, dark energy evolves more and more slowly as time progresses; in the thawing variant, the field changes slowly at first and faster later.

A third option may account for cosmic acceleration: there is no dark energy, and the quickening expansion of the universe results from physics not explained by Einstein's theory of gravity (general relativity), which is incomplete. It is possible that in truly extreme regimes, such as the breadth of galaxy clusters or the entire observable universe, the laws of gravity work differently than the theory predicts, and gravity misbehaves. Physicists have put forth a few interesting theoretical suggestions along these lines, but no self-consistent theory that agrees with all the observations currently exists, so dark energy seems to have the upper hand over this option for now. (Previous ideas, such as the notion that cosmic acceleration is a manifestation of an uneven distribution of matter throughout the universe or the result of a network of geometric defects in the structure of space, have by now largely proved to be inconsistent with observational data.)

WHY IS DARK ENERGY SO WEAK?

NONE OF THE PROPOSED explanations for dark energy is very satisfying. The cosmological constant, for example, predicts that dark energy should be vastly stronger than it actually is. When one naively attempts to sum up the energies over all the presumed quantum states associated with the sea of virtual particles and antiparticles in the vacuum of space, one obtains a value that is more than 120 orders of magnitude larger than the observed sum. Factoring in ideas from proposed theories such as supersymmetry—the notion that every known particle has a heavier partner particle that we have not yet discovered—reduces the discrepancy somewhat, but the difference between the predicted and the measured total energy still remains tens of orders of magnitude too high. So if dark energy is explained by the energy of the vacuum, the question is, How did this vacuum energy turn out to be so tiny?

The field explanation for dark energy barely does better on this front. Theorists simply assume (without a good explanation for why it should be so) that the minimum of the potential energy associated with the dark energy field is very low, guaranteeing that only a small amount of dark energy is spread throughout space. Also, these models require the field to have surprisingly minimal interactions with everything else in the universe (apart from its

IN BRIEF

Almost 20 years ago scientists discovered that the expansion of the universe is accelerating and dubbed the source of this acceleration "dark energy."
Intensive investigations since then have not resolved the nature of dark energy but rather exposed a number of further questions: Why is dark energy so much weaker than theory seems to predict, what does it mean for the future of the cosmos, and might dark energy lead to the conclusion that we live in a multiverse?
With the advent of several recent and upcoming experiments, scientists hope that the coming years will finally bring some answers.

© 2017 Scientific American

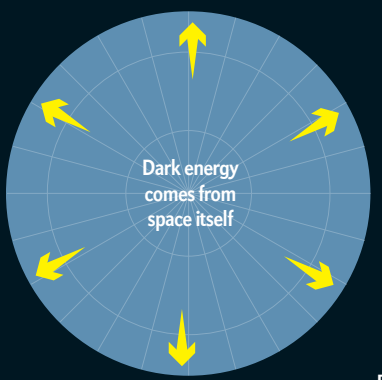
Dark Energy Possibilities and Potential Futures

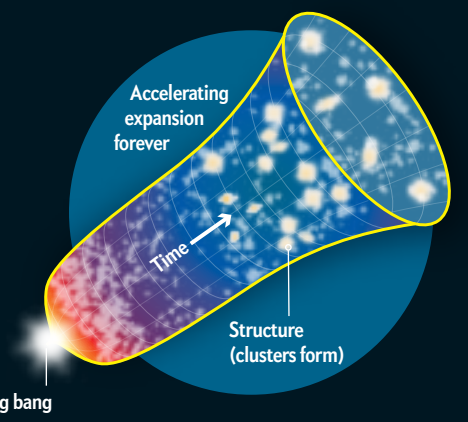
Dark energy is scientists' name for whatever is causing the expansion of the universe to accelerate. Explanations for dark energy fall into three main categories: it may be an unchanging energy arising from empty space (an idea called the cosmological constant) or a varying energy stemming from a field pervading the universe (quintessence). Or dark energy may not exist at all—in that case, gravity would act differently than thought on cosmic scales.

MODEL ····▶ FUTURE

Cosmological Constant

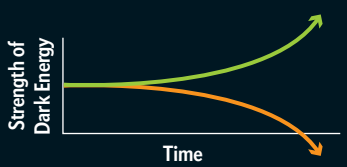
If the vacuum of empty space has an inherent energy, it may push the universe to expand. The strength of such an energy would be constant through time and would act just like the cosmological constant term Albert Einstein added to, and later removed from, his equations of general relativity.

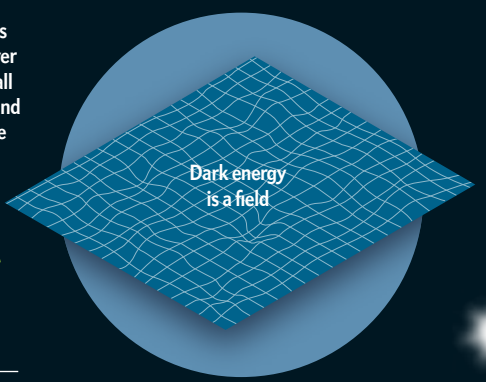


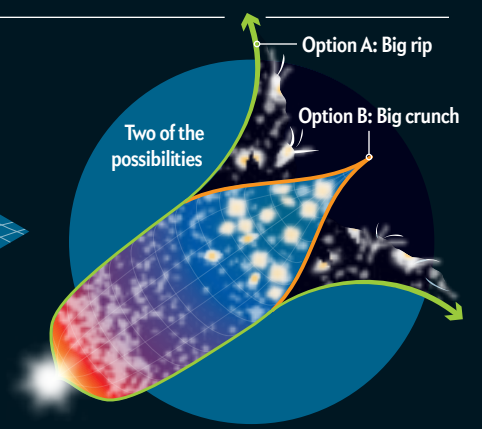


Quintessence

If dark energy comes from a field that fills the cosmos, its strength could change over time, either increasing to eventually rip all structures in space apart or decreasing and changing directions to allow the universe to contract in a big crunch.







There Is No Dark Energy

Dark energy may not exist at all, and the acceleration of the universe's expansion may instead indicate that gravity operates differently than we think on extremely large scales.

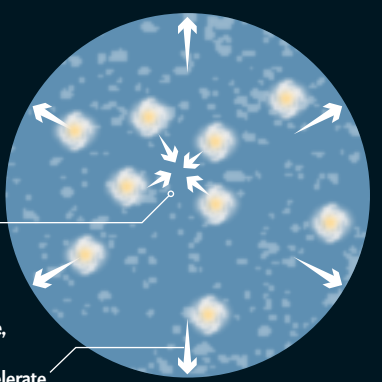
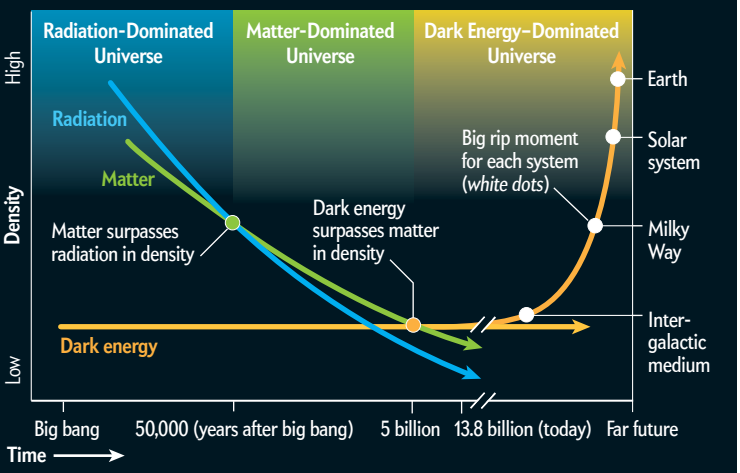

On the scale of galaxies and clusters, gravity behaves as general relativity predicts
On the scale of the universe as a whole, gravity grossly diverges from general relativity; the universe appears to accelerate

Illustration by Nigel Hawtin

© 2017 Scientific American

Cosmic Control

Because dark energy is denser in space than any other constituent of the universe, it exerts the dominant influence on the cosmos and will therefore control its fate. Dark energy was not always on top, though: the other ingredients of the universe—radiation (light) and matter (including atoms and regular matter as well as invisible dark matter)—were dominant when the universe was young and small, and they were packed tightly in space. As the universe expanded over time, matter and radiation spread out, and dark energy overpowered them. If the density of dark energy increases, it may become so powerful that it rips apart all structures in space.



repulsive gravitational push)—a property that is hard to explain. These facts make it difficult to incorporate the dark energy field hypotheses naturally within common particle physics models.

WHAT DOES THIS MEAN FOR THE FUTURE OF THE UNIVERSE?

THE PROPERTIES of dark energy will determine the ultimate fate of our universe. For instance, if dark energy is indeed the energy of empty space (the cosmological constant), the acceleration will continue forever, and about a trillion years from now expansion will cause all the galaxies that are more distant than the Milky Way's closest neighbors (the Local Group, which by then will coalesce to form one large elliptical galaxy) to separate faster than light speed, rendering them undetectable. Even the ancient light from the afterglow of the big bang—the cosmic microwave background (CMB) that fills all of space—will be stretched to wavelengths longer than the size of the visible universe and thus rendered imperceptible. In this scenario, we happen to be living at a very fortunate time when we still have the best possible view of our universe.

If, on the other hand, dark energy is not the energy of the vacuum but rather the energy of some unknown field, then the future is wide open. Depending on the way the field evolves, the universe could eventually stop expanding and start collapsing, falling in on itself to a final "big crunch" that mimics the big bang from whence it came. Or the universe could end up in a "big rip," in which all complex structures, from clusters of galaxies to atoms and atomic nuclei, could become overwhelmed by dark energy and tear apart. And the first scenario, continuous acceleration toward a cold death, is also an option with a dark energy field.

An alternative theory of gravity, if it turns out to be necessary, likewise allows for various outcomes depending on the particularities of the revised theory.

MIGHT WE LIVE IN A MULTIVERSE?

WITH THE COSMOLOGICAL CONSTANT explanation leading the theoretical pack, the problem of its inexplicable weakness comes to the forefront. Realizing this problem with the constant even before the discovery of accelerating expansion, physicist Steven Weinberg of the University of Texas at Austin suggested a new paradigm—one in which the cosmological constant is not uniquely determined from the basic laws of physics but rather is a random variable that assumes different values in different members of a huge ensemble of universes—a multiverse. Some universes may have much larger cosmological constants, but in those the accelerating repulsive force is so large that matter cannot coalesce to form galaxies, planets and life. Because we exist, Weinberg reasons, we clearly have to find ourselves in one of those universes that *can* allow for our existence—one that happened to have a small cosmological constant. This idea, which was further developed by Alexander Vilenkin of Tufts University, Martin Rees of the University of Cambridge, one of us (Livio) and others, is called anthropic reasoning.

There are good reasons, aside from the dark energy consideration, that a multiverse might arise. The widely accepted theory of cosmic inflation suggests the universe ballooned stupendously in its first fraction of a second. Vilenkin and Andrei Linde of Stanford University have shown that once cosmic inflation starts, it is essentially impossible to stop it from occurring again and again, thereby creating an infinite ensemble of bubbles or "pocket universes" that form in isolation from one another and may have very different properties.

A multiverse also seems to be a consequence of string theory, a candidate for a theory that unifies all the forces of nature. Calculations based on versions of string theory called M-theory by Raphael Bousso and Joseph Polchinski suggest that there could be as many as 10^{500} different spacetimes, or universes, each characterized by different values for the constants of nature and even the number of dimensions in space.

Even just mentioning the multiverse idea, however, raises the blood pressure of some physicists. The notion seems hard to swallow and harder to test—perhaps signifying the end of the classical scientific method as we know it. Historically this method has required that hypotheses should be directly testable by new experiments or observations. Yet the concept of the multiverse does make a few predictions that may yield to testing. In particular,

Graphic by Jen Christiansen

© 2017 Scientific American

some multiverse models predict that the shape of spacetime has a slight curve to it that might be detectable by observations. Another possibility, albeit not a very likely one, is that the cosmic microwave background light may contain ripples that are the signature of a collision of another bubble with our own.

FINDING ANSWERS

THE BEST WAY we know of to start to reveal the nature of dark energy is to measure the ratio of its pressure (how much it is in tension with space) to its density (how much of it is in a given amount of space)—a property called its equation of state parameter, w. If dark energy is the energy of the vacuum (the cosmological constant), then w will be constant and equal to –1. If dark energy is associated with a field that changes with time, on the other hand, we would hope to detect its w value differing from –1 and evolving throughout cosmic history. Alternatively, if the observed acceleration necessitates a modification of Einstein's theory of gravity for extreme distances, we would hope to observe an inconsistency between the value of w we find at different scales in the universe.

Astronomers have devised some clever indirect ways to measure dark energy's pressure and density. As a repulsive gravitational pull, dark energy or modified gravity would counteract the pull of regular gravity (which tugs mass in the universe toward other mass), discouraging the formation of large-scale structure—that is, galaxy clusters. Thus, by studying how clusters grew over time, scientists can find out how strong dark energy was at various points in history. We do this by observing how the mass of clusters bends the light of background galaxies behind them through a process known as gravitational lensing. The amount of bending tells us how massive the clusters are, and by observing this effect for clusters at varying distances, we can measure how common massive clusters were at various cosmic epochs. (Because of the time it takes light to reach us, looking far away is akin to looking back in time.)

We can also measure dark energy by studying how the rate of the universe's expansion has changed over time. By observing objects at different distances and measuring their redshift—how much the wavelength of their light has been stretched by the expansion of space—we can learn how much the universe has expanded since that light began its journey. This method, in fact, was how the two teams initially discovered cosmic acceleration; they measured the redshifts of different type 1a supernovae (whose distances are reliably tied to their brightness). A variation on this technique is to observe the apparent size of ripples in the density of galaxies across space called baryon acoustic oscillations (BAOs)—another reliable distance indicator—as a way to trace the expansion history of the universe.

To date, most measurements of w are generally consistent (within the observational uncertainties) with a value of –1 to within 10 percent and thus support the cosmological constant explanation of accelerated expansion. Recently a team led by Riess used the Hubble Space Telescope to probe dark energy back to about 10 billion years ago using the supernova technique and found no evidence for variation over that time.

It is worth noting, however, that some hints of a deviation from cosmological constant predictions have shown up in the past couple of years. For instance, a combination of measurements of the CMB (which tells us about the total mass and energy in the universe) from the Planck satellite with results from gravitational-lensing studies suggests a value for w that is more negative than –1. Observations from the first Panoramic Survey Telescope and Rapid Response System (Pan-STARRS), using more than 300 supernovae to track cosmic expansion, also seemed to indicate a value of w more negative than –1. Very recent observations of baryon acoustic oscillations in data from distant bright galaxies called quasars seem to hint that dark energy density has increased with time. Finally, a small discrepancy between local measurements of the rate of space's expansion today, compared with measurements of the primordial rate of expansion from the CMB, could also be pointing to a deviation from a cosmological constant. As intriguing as these results are, none is yet compelling. More data in the near future may strengthen these discrepancies or reveal them to be systematic flukes.

Work is now under way to achieve a 100-fold improvement in the precision of the measured properties of dark energy in the coming decade. New projects such as the Dark Energy Survey (DES) begun in 2013 and the Large Synoptic Survey Telescope (LSST) due to open around 2021 will gather better information about large-scale structure in the universe and the history of expansion. NASA's Wide-Field Infrared Survey Telescope–Astrophysics Focused Telescope Assets (WFIRST-AFTA), a 2.4-meter space telescope planned to launch in the mid-2020s, is expected to observe distant supernovae and BAOs, as well as gravitational lensing. The launch of the European Space Agency's Euclid space mission is currently planned for 2020 and will also exploit lensing, BAOs and redshift measurements of galactic distances to determine the three-dimensional distribution of galaxy clusters.

Finally, we can also test theories of modified gravity through measurements within the solar system. One method measures the distance to the moon to such an astonishingly high precision (through reflecting laser light off reflectors placed on the moon by *Apollo* astronauts) that it can detect minute deviations from predictions of general relativity. In addition, ingenious laboratory experiments will search for minuscule discrepancies in the current laws of gravity.

The coming years should be a pivotal time for dark energy research. We have hopes of being able to make real progress in addressing the outstanding questions about the accelerating expansion of the universe. The answers will reveal nothing short of the future of the cosmos. SA

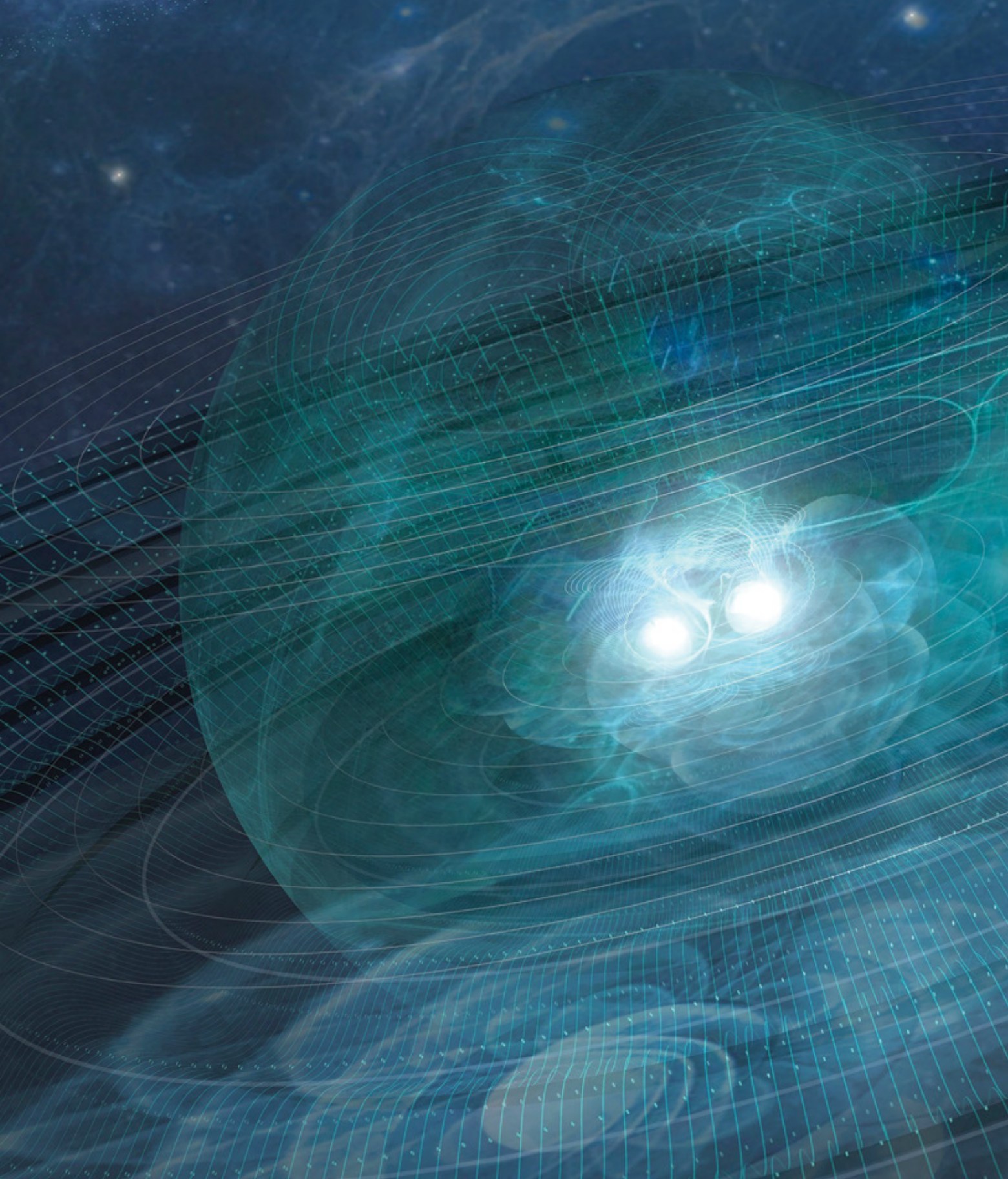
Adam G. Riess is an astrophysicist at Johns Hopkins University and the Space Telescope Science Institute. His research on distant supernovae revealed that the expansion of the universe is accelerating, a discovery for which he shared the 2011 Nobel Prize in Physics.
Mario Livio is an astrophysicist who worked for 24 years with the Hubble Space Telescope and a best-selling author of popular science books; his most recent one is *Why?: An Exploration of Human Curiosity* (Simon & Schuster, 2017).

MORE TO EXPLORE

Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. Adam G. Riess et al. in *Astronomical Journal,* Vol. 116, No. 3, pages 1009–1038; September 1998.
The Accelerating Universe. Mario Livio. Wiley, 2000.

scientificamerican.com/magazine/sa

© 2017 Scientific American



RADIATING RIPPLES created by two black holes colliding are seen in this artist's rendition.

© 2017 Scientific American

RIPPLES IN SPACETIME

A daring set of experiments finally confirms Einstein's 1916 prediction of gravitational waves

By Clara Moskowitz

ON FEBRUARY 11, 2016, SCIENTISTS ANNOUNCED THAT THE Laser Interferometer Gravitational-Wave Observatory (LIGO) had discovered ripples in spacetime created by merging black holes. About 1.3 billion years ago two black holes swirled closer and closer together until they crashed in a furious bang. Each black hole packed roughly 30 times the mass of our sun into a minute volume, and their head-on impact came as the two were approaching the speed of light. The staggering strength of the merger gave rise to a new black hole and created a gravitational field so strong that it distorted spacetime in waves that spread throughout space with a power about 50 times stronger than that of all the shining stars and galaxies in the observable universe.

Illustration by Moonrunner Design

© 2017 Scientific American

Such events are, incredibly, thought to be common in space, but this collision was the first of its kind ever detected and its waves the first ever seen. At the much anticipated LIGO conference in Washington, D.C. (one of at least five simultaneous events held in the U.S. and Europe), the scientists announced that the more than half-century search for gravitational waves had finally succeeded.

"This was truly a scientific moonshot, and we did it, we landed on the moon," LIGO executive director David Reitze said during the announcement.

"There are people who've put their entire life into this search, and there are people who died before having a chance to see anything," says LIGO team member Szabolcs Márka, a physicist at Columbia University. "It's really a wonderful feeling that you have validated the investment of the tremendous amount of work. And it's not just that you found something, but you gave something to everybody, to the rest of the human race."

"IT'S NOT JUST THAT YOU FOUND SOMETHING, BUT YOU GAVE SOMETHING TO EVERYBODY, TO THE REST OF THE HUMAN RACE."

—SZABOLCS MÁRKA, *Columbia University*

Albert Einstein first predicted gravitational waves in 1916 based on his general theory of relativity, but even he waffled about whether or not they truly exist. Scientists began seeking these ripples in spacetime in the 1960s, but none succeeded in measuring their effects on Earth until now. LIGO's first discovery, accepted for publication in *Physical Review Letters,* not only provided the first direct evidence for gravitational waves but also opens the door to using them to study the powerful cosmic events that create them. "It's a huge deal," says Luis Lehner, a physicist at the Perimeter Institute for Theoretical Physics in Ontario, who is not affiliated with the LIGO project. "It has pushed the fundamental theory of gravity forward in a very strong way and gives us an incredible tool to probe very deep questions of the universe."

After the initial observation, the LIGO team announced two more discoveries of gravitational waves, including a signal presented later in 2016 created by two black holes that were smaller in mass than the original pair and another signal reported this past spring from two black holes about three times as distant as the first collision.

"ONE OF THE MOST COMPLEX SYSTEMS EVER BUILT"

MORE THAN 1,000 SCIENTISTS work on the $1-billion LIGO experiment, which is funded by the National Science Foundation. The project uses two detectors, one located in Washington State and the other in Louisiana, to sense the distortions in space that occur when a gravitational wave passes through Earth. Each detector is shaped like a giant L, with legs four kilometers long. Laser light bounces back and forth through the legs, reflecting off mirrors, and amazingly precise atomic clocks measure how long it takes to make the journey. Typically the two legs are exactly the same length, and so the light takes exactly the same amount of time to traverse each. If a gravitational wave passes through, however, the detector and the ground underneath it will expand and contract infinitesimally in one direction, and the two perpendicular legs will no longer be the same size. One of the lasers will arrive a fraction of a second later than the other.

LIGO must be unbelievably sensitive to measure this change in the length of the legs, which is smaller than one ten-thousandth the diameter of a proton, or less than the size of a soccer ball compared with the span of the Milky Way. "It's one of the most complex systems ever built by mankind," Márka says. "There are so many knobs to turn, so many things to align, to achieve that [sensitivity]." In fact, the experiment is so delicate that unrelated events such as an airplane flying overhead, wind buffeting the building or tiny seismic shifts in the ground below the detector can disturb the lasers in ways that mimic gravitational signals. "If I clap in the control room, you will see a blip," says Imre Bartos, another member of the LIGO team at Columbia. The researchers carefully weed out such contaminating signals and also take advantage of the fact that the detectors in Washington and Louisiana are highly unlikely to be affected by the same contamination at the same time. "By comparing the two detectors, we can be even more certain that what we are seeing is something that's coming from outside Earth," Bartos adds.

LIGO began its first run in 2002 and hunted through 2010 without finding any gravitational waves. The scientists then shut down the experiment and upgraded nearly every aspect of the detectors, including boosting the power of the lasers and replacing the mirrors, for a subsequent run, called Advanced LIGO, that began officially on September 18, 2015. Yet even before then, the experiment was up and running: the signal arrived on September 14 at 5:51 A.M. Eastern time, reaching the detector in Louisiana seven milliseconds before it got to the detector in Washington. Advanced LIGO is already about three times more sensitive than the initial LIGO and is designed to become about 10 times more sensitive than the first iteration in the next few years.

LONG TIME COMING

BEFORE THIS DISCOVERY, the strongest evidence of gravitational waves came indirectly from observations of superdense,

IN BRIEF

On February 11, 2016, scientists at the Laser Interferometer Gravitational-Wave Observatory (LIGO) announced that they had successfully detected gravitational waves for the first time in history.

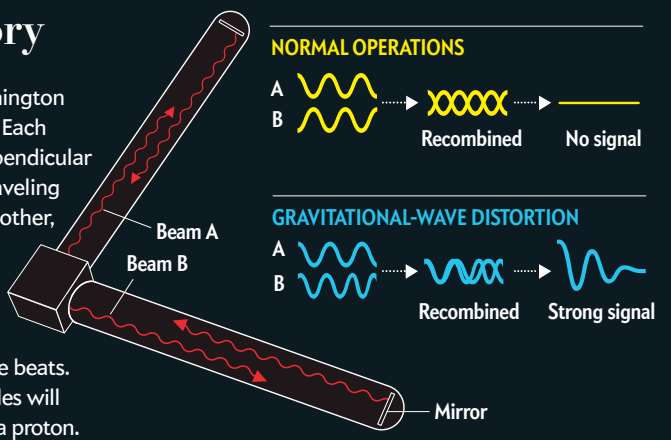
Before this discovery, evidence for the existence of gravitational waves had come indirectly from observations of pulsars. LIGO's detection is also the strongest confirmation to date for the existence of black holes.

LIGO will enable scientists to study black holes and other cosmic events, such as supernovae and neutron stars, in entirely new ways.

© 2017 Scientific American

The Laser Interferometer Gravitational-Wave Observatory

LIGO experiment uses two L-shaped detectors, one in Washington State and one in Louisiana, to search for gravitational waves. Each detector bounces laser light between mirrors down two perpendicular four-kilometer legs. LIGO splits the light so that the wave traveling through the first leg, beam A, is out of phase with that in the other, beam B. When the beams recombine (*yellow*), the waves should cancel each other out, rendering the resulting beam dark. If, however, a gravitational wave passes through Earth and changes the relative length of the legs (*blue*), the waves will not match up and the combined beams will reveal telltale beats. The effect is tiny, however—a nearby collision of two black holes will change the length of LIGO's legs by less than the diameter of a proton.



spinning neutron stars called pulsars. In 1974 Joseph Taylor, Jr., and Russell Hulse discovered a pulsar circling a neutron star, and later observations showed that the pulsar's orbit was shrinking. Scientists concluded that the pulsar must be losing energy in the form of gravitational waves—a discovery that won Taylor and Hulse the 1993 Nobel Prize in Physics. Ever since this clue, astronomers have been hoping to detect the waves themselves. "I've certainly been looking forward to this event for a long time," Taylor says. "There is a long history, and I think projects that take this long to bear fruit require an awful lot of patience. It's about time."

The discovery is not just proof of gravitational waves but the strongest confirmation yet for the existence of black holes. "We think black holes exist out there. We have very strong evidence they do, but we don't have direct evidence," Lehner says. "Everything is indirect. Given that black holes themselves cannot give any signal other than gravitational waves, this is the most direct way to prove that a black hole exists."

LIGO's ability to study the characteristics of gravitational waves will allow scientists to study black holes in a whole new way. Researchers would like to know the details of how two black holes collide and whether a new black hole arises as theory suggests. "We're talking about two objects that do not emit light—they're completely dark," says Janna Levin, a theorist at Barnard College, who is outside the LIGO collaboration. "In the details of a collision and in terms of the gravitational waves, you could see the formation of a new black hole." The observatory should also be able to detect gravitational waves created by other cataclysmic events, such as exploding supernovae and collisions of two neutron stars.

LIGO and future gravitational-wave experiments will also allow physicists to put general relativity to the test. The 100-year-old theory has stood the test of time, but it still conflicts with the theory of quantum mechanics that rules over the subatomic realm. "We know general relativity should show cracks at some point, and the way it shows them will guide our theory to one that is more complete," Lehner says. "This is pushing the theory over six orders of magnitude compared with the previous strongest test," which came from observations of pulsars.

Advanced LIGO is now on its second observing run and will shut down for upgrades over a period of about 15 months starting later this year. Scientists expect to increase LIGO's sensitivity by a factor of four or five times, which should allow it to detect black hole binaries about 100 times as often as it currently does. LIGO is the first of many observatories that will join this new era of gravitational astronomy. A similar project, Italy's Advanced Virgo experiment, has joined with LIGO, although it is running at a lower sensitivity. Later this decade the Kamioka Gravitational Wave Detector (KAGRA) in Japan will begin observations. And plans for a detector in India are in the works. Ground-based telescope projects called pulsar timing arrays aim to study gravitational waves by noting delays in light from pulsars arriving on Earth after traveling through wave-stretched space. And a spacecraft called Lisa Pathfinder concluded in June 2017 its 16-month mission to test technology for a proposed space-based observatory that will be sensitive to longer-wavelength gravitational waves from supermassive black hole collisions.

"Every time you open a new window to the universe we always discover new things," Lehner says. "It's like Galileo pointing the first telescope to the sky. Initially he saw some planets and moons, but then as we got radio, UV and x-ray telescopes, we discovered more and more about the universe. We are pretty much at the moment where Galileo was beginning to see the first objects around Earth. It will have such a huge impact on the field." SA

Clara Moskowitz is *Scientific American*'s senior editor covering space and physics. She has a bachelor's degree in astronomy and physics from Wesleyan University and a graduate degree in science journalism from the University of California, Santa Cruz.

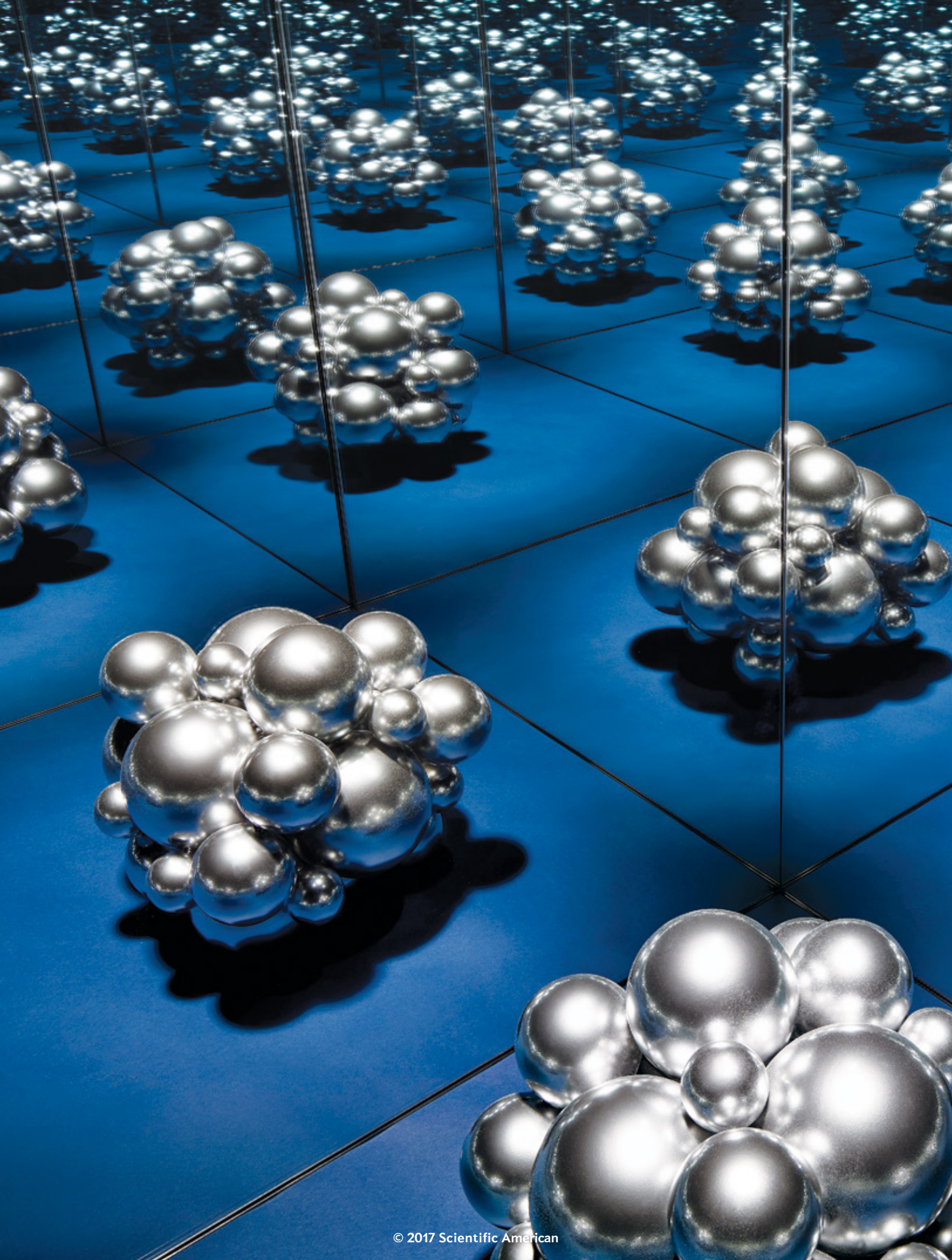
MORE TO EXPLORE

Observation of Gravitational Waves from a Binary Black Hole Merger. B. P. Abbott et al. in *Physical Review Letters*, Vol. 116, Article 061102; February 11, 2016.
The Next Wave. Davide Castelvecchi in *Nature*, Vol. 531, pages 428–431; March 24, 2016.

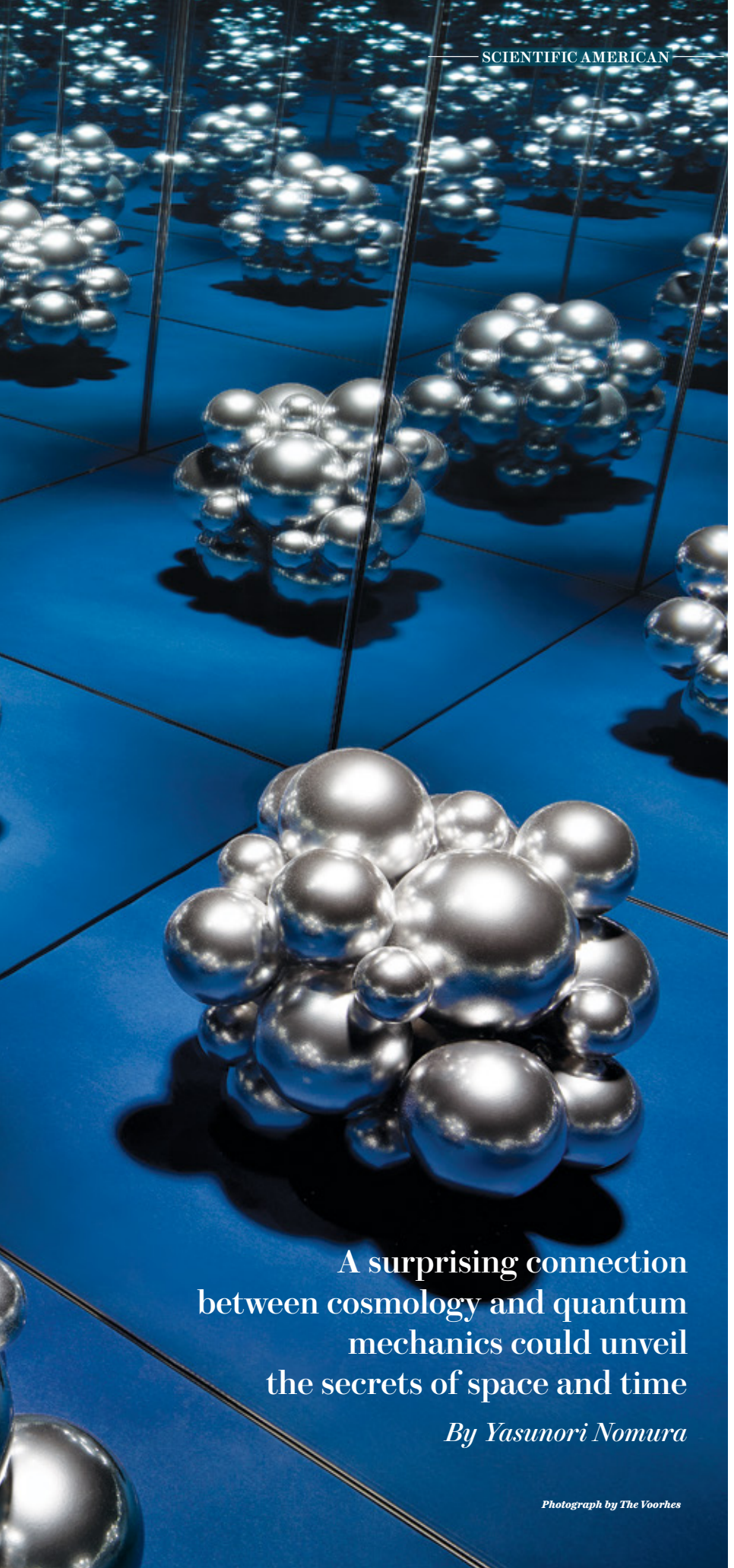
scientificamerican.com/magazine/sa

Graphic by Jen Christiansen

© 2017 Scientific American



© 2017 Scientific American



THE QUANTUM MULTIVERSE

A surprising connection between cosmology and quantum mechanics could unveil the secrets of space and time

By Yasunori Nomura

Photograph by The Voorhes

© 2017 Scientific American

MANY COSMOLOGISTS NOW ACCEPT THE EXTRAORDINARY IDEA THAT what seems to be the entire universe may actually be only a tiny part of a much larger structure called the multiverse. In this picture, multiple universes exist, and the rules we once assumed were basic laws of nature take different forms in each; for example, the types and properties of elementary particles may differ from one universe to another. The multiverse idea emerges from a theory that suggests the very early cosmos expanded exponentially. During this period of "inflation," some regions would have halted their rapid expansion sooner than others, forming what are called bubble universes, much like bubbles in boiling water. Our universe would be just one of these bubbles, and beyond it would lie infinitely more.

The idea that our entire universe is only a part of a much larger structure is, by itself, not as outlandish as it sounds. Throughout history scientists have learned many times over that the visible world is far from all there is. Yet the multiverse notion, with its unlimited number of bubble universes, does present a major theoretical problem: it seems to erase the ability of the theory to make predictions—a central requirement of any useful theory. In the words of Alan Guth of the Massachusetts Institute of Technology, one of the creators of inflation theory, "in an eternally inflating universe, anything that can happen will happen; in fact, it will happen an infinite number of times."

In a single universe where events occur a finite number of times, scientists can calculate the relative probability of one event occurring versus another by comparing the number of times these events happen. Yet in a multiverse where everything happens an infinite number of times, such counting is not possible, and nothing is more likely to occur than anything else. One can make any prediction one wants, and it is bound to come true in some universe, but that fact tells you nothing about what will go on in our specific world.

This apparent loss of predictive power has long troubled physicists. Some researchers, including me, have now realized that quantum theory—which, in contrast to the multiverse notion, is concerned with the very smallest particles in existence—may, ironically, point the way to a solution. Specifically, the cosmological picture of the eternally inflating multiverse may be mathematically equivalent to the "many worlds" interpretation of quantum mechanics, which attempts to explain how particles can seem to be in many places at once. As we will see, such a connection between the theories not only solves the prediction problem, it may also reveal surprising truths about space and time.

QUANTUM MANY WORLDS

I CAME TO THE IDEA of a correspondence between the two theories after I revisited the tenets of the many-worlds interpretation of quantum mechanics. This concept arose to make sense of some of the stranger aspects of quantum physics. In the quantum world—a nonintuitive place—cause and effect work differently than they do in the macro world, and the outcome of any process is always probabilistic. Whereas in our macroscopic experience, we can predict where a ball will land when it is thrown based on its starting point, speed and other factors, if that ball were a quantum particle, we could only ever say it has a certain chance of ending up here and another chance of ending up there. This probabilistic nature cannot be avoided by knowing more about the ball, the air currents or such details; it

IN BRIEF

The theory of cosmic inflation, which implies that the early cosmos expanded exponentially, suggests that we live not in a universe but a vast multiverse.
The problem with the multiverse idea, however, is that all events that can occur will occur infinitely many times, ruining the theory's predictive ability.
Physicists realized they can resolve the issue by viewing the multiverse as equivalent to a notion from quantum mechanics called the many-worlds interpretation, which suggests that our universe is one of many that coexist in "probability space" rather than in a single real space.

© 2017 Scientific American

COURTESY OF NASA, ESA, S. BECKWITH (*STScI*) AND HUBBLE DEEP FIELD TEAM

HUBBLE SPACE TELESCOPE'S Ultra Deep Field shows galaxies as far away as 13 billion light-years. Objects much farther out will forever be beyond reach because the expansion of space causes them to recede faster than the speed of light. This so-called cosmological horizon has important implications for the theory of the multiverse.

is an intrinsic property of the quantum realm. The same exact ball thrown under the same exact conditions will sometimes land at point A and other times at point B. This conclusion may seem strange, but the laws of quantum mechanics have been confirmed by innumerable experiments and truly describe how nature works at the scale of subatomic particles and forces.

In the quantum world, we say that after the ball is thrown, but before we look for its landing spot, it is in a so-called superposition state of outcomes A and B—that is, it is neither at point A nor point B but located in a probabilistic haze of *both* points A and B (and many other locations as well). Once we look, however, and find the ball in a certain place—say, point A—then anyone else who examines the ball will also confirm that it sits at A. In other words, before any quantum system is measured, its outcome is uncertain, but afterward all subsequent measurements will find the same result as the first.

In the conventional understanding of quantum mechanics, called the Copenhagen interpretation, scientists explain this shift by saying that the first measurement changed the state of the system from a superposition state to the state A. But although the Copenhagen interpretation does predict the outcomes of laboratory experiments, it leads to serious difficulties at the conceptual level. What does the "measurement" really mean, and why does it change the state of the system from a superposition of possibilities to a single certainty? Does the change of state occur when a dog or even a fly observes the system? What about when a molecule in the air interacts with the system, which we expect to be occurring all the time yet which we do not usually treat as a measurement that can interfere with the outcome? Or is there some special physical significance in a human consciously learning the state of the system?

In 1957 Hugh Everett, then a graduate student at Princeton University, developed the many-worlds interpretation of quantum mechanics that beautifully addresses this issue—although at the time many received it with ridicule, and the idea is still less favored than the Copenhagen interpretation. Everett's key insight was that the state of a quantum system reflects the state of the *whole* universe around it, so that we must include the observer in a complete description of the measurement. In other words, we cannot consider the ball, the wind and the hand that throws it in isolation—we must also include in the fundamental description the person who comes along to inspect its landing spot, as well as everything else in the cosmos at that time. In this picture, the quantum state after the measurement is still a superposition—not just a superposition of two landing spots but of two entire worlds! In the first world, the observer finds that the state of the system has changed to A, and therefore any observer in this particular world will obtain result A in all subsequent measurements. But when the measurement was made, another universe split off from the first in which

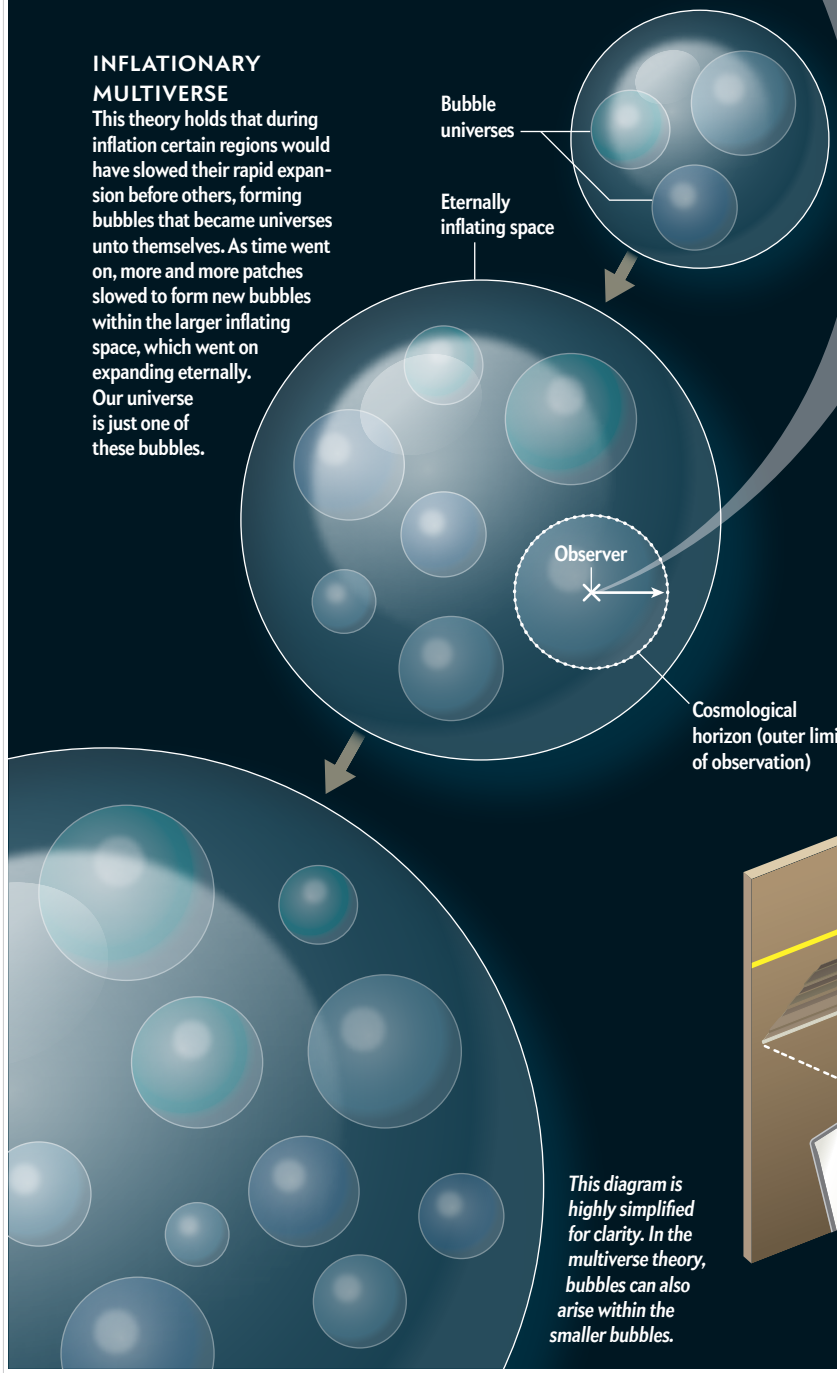
© 2017 Scientific American

Inflation Meets Many Worlds

The theory of inflation suggests that our universe is one of infinitely many that formed when the very early cosmos expanded exponentially. This picture of a multiverse, however, seems to destroy the theory's ability to make predictions because anything that can happen in an infinite multiverse will happen infinitely many times. The problem is solved, however, if the inflationary multiverse is equivalent to the "many worlds" interpretation of quantum mechanics, which posits that all these infinite universes coexist not in a single real space but in "probability space."

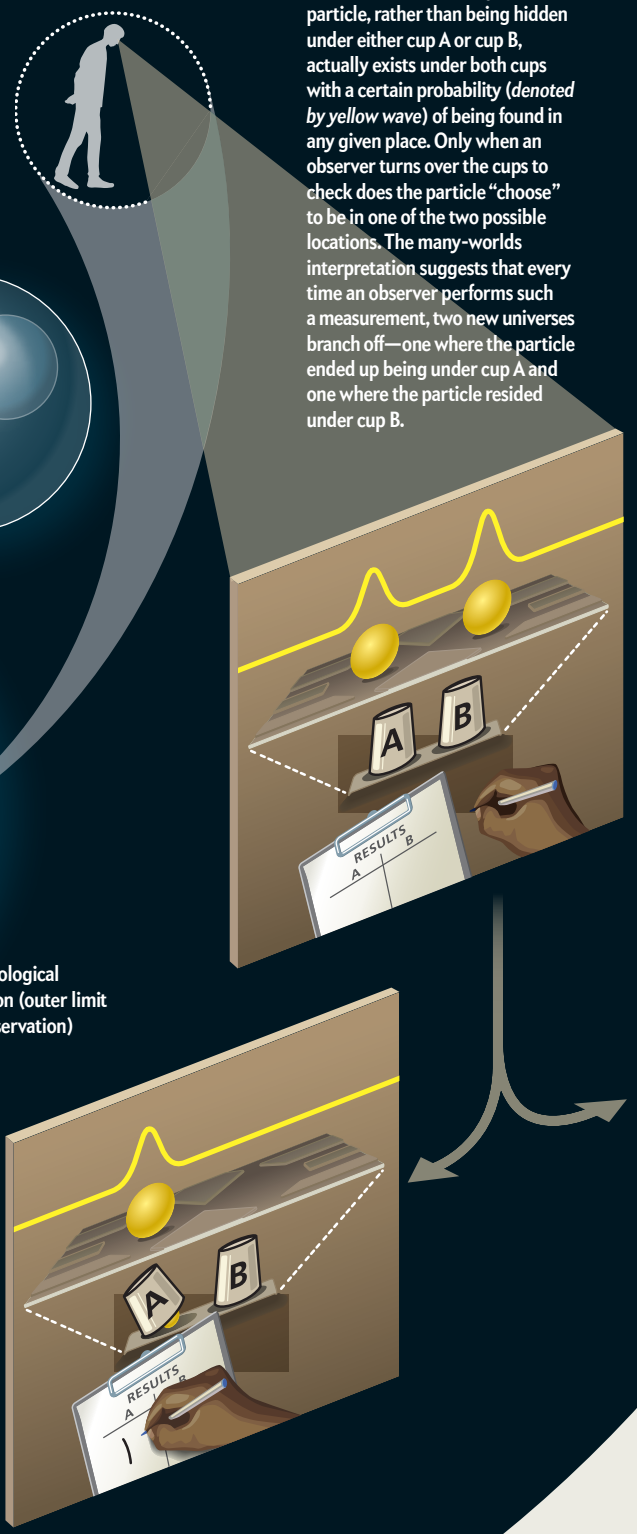
INFLATIONARY MULTIVERSE

This theory holds that during inflation certain regions would have slowed their rapid expansion before others, forming bubbles that became universes unto themselves. As time went on, more and more patches slowed to form new bubbles within the larger inflating space, which went on expanding eternally. Our universe is just one of these bubbles.



MANY WORLDS

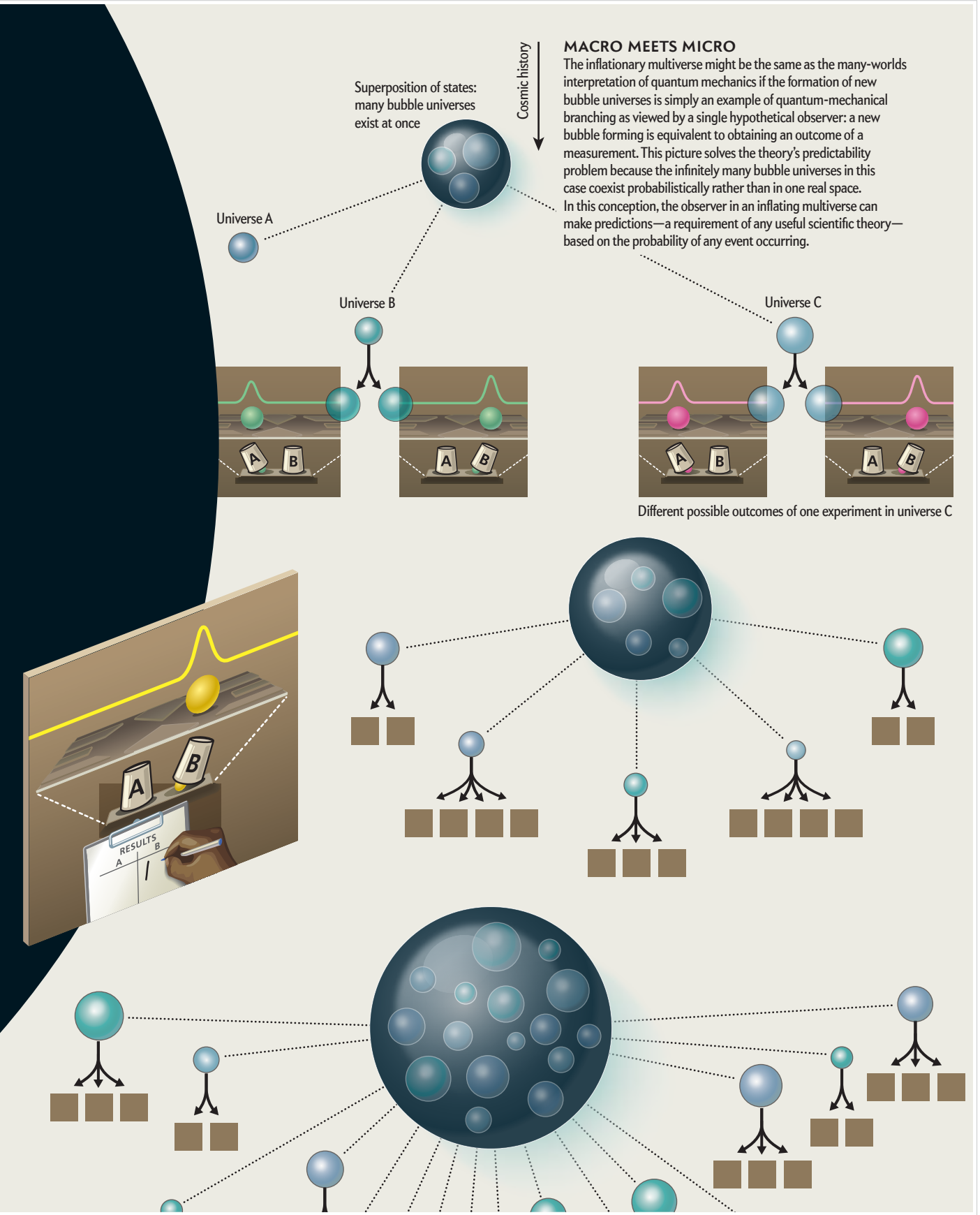
Quantum mechanics says that a particle, rather than being hidden under either cup A or cup B, actually exists under both cups with a certain probability (*denoted by yellow wave*) of being found in any given place. Only when an observer turns over the cups to check does the particle "choose" to be in one of the two possible locations. The many-worlds interpretation suggests that every time an observer performs such a measurement, two new universes branch off—one where the particle ended up being under cup A and one where the particle resided under cup B.



This diagram is highly simplified for clarity. In the multiverse theory, bubbles can also arise within the smaller bubbles.

Illustration by Jen Christiansen

© 2017 Scientific American



MACRO MEETS MICRO

The inflationary multiverse might be the same as the many-worlds interpretation of quantum mechanics if the formation of new bubble universes is simply an example of quantum-mechanical branching as viewed by a single hypothetical observer: a new bubble forming is equivalent to obtaining an outcome of a measurement. This picture solves the theory's predictability problem because the infinitely many bubble universes in this case coexist probabilistically rather than in one real space. In this conception, the observer in an inflating multiverse can make predictions—a requirement of any useful scientific theory—based on the probability of any event occurring.

© 2017 Scientific American

the observer finds, and keeps finding, that the ball landed at point B. This feature explains why the observer—let us say it is a man—thinks that his measurement changes the state of the system; what actually happens is that when he makes a measurement (interacts with the system), he himself divides into two different people who live in two different parallel worlds corresponding to two separate outcomes, A and B.

According to this picture, humans making measurements have no special significance. The state of the entire world continuously branches into many possible parallel worlds that coexist as a superposition. A human observer, being a part of nature, cannot escape from this cycle—the observer keeps splitting into many observers living in many possible parallel worlds, and all are equally "real." An obvious but important implication of this picture is that everything in nature obeys the laws of quantum mechanics, whether small or large.

I AND OTHER PHYSICISTS ARE ALSO PURSUING THE QUANTUM MULTIVERSE IDEA FURTHER. HOW CAN WE DETERMINE THE QUANTUM STATE OF THE ENTIRE MULTIVERSE? WHAT IS TIME, AND HOW DOES IT EMERGE?

What does this interpretation of quantum mechanics have to do with the multiverse discussed earlier, which seems to exist in a continuous real space rather than as parallel realities? In 2011 I argued that the eternally inflating multiverse and quantum-mechanical many worlds à la Everett are the same concept in a specific sense. In this understanding, the infinitely large space associated with eternal inflation is a kind of "illusion"—the many bubble universes of inflation do not all exist in a single real space but represent the possible different branches on the probabilistic tree. Around the same time that I made this proposal, Raphael Bousso of the University of California, Berkeley, and Leonard Susskind of Stanford University put forth a similar idea. If true, the many-worlds interpretation of the multiverse would mean that the laws of quantum mechanics do not operate solely in the microscopic realm—they also play a crucial role in determining the global structure of the multiverse even at the largest distance scales.

BLACK HOLE QUANDARY

TO BETTER EXPLAIN how the many-worlds interpretation of quantum mechanics could describe the inflationary multiverse, I must digress briefly to talk about black holes. Black holes are extreme warps in spacetime whose powerful gravity prevents objects that fall into them from escaping. As such, they provide an ideal testing ground for physics involving strong quantum and gravitational effects. A particular thought experiment about these entities reveals where the traditional way of thinking about the multiverse goes off track, thereby making prediction impossible.

Suppose we drop a book into a black hole and observe from the outside what happens. Whereas the book itself can never escape the black hole, theory predicts that the information in the book will not be lost. After the book has been shredded by the black hole's gravity and after the black hole itself has gradually evaporated by emitting faint radiation (a phenomenon known as Hawking radiation, discovered by physicist Stephen Hawking of the University of Cambridge), outside observers can reconstruct all the information contained in the initial book by closely examining the radiation released. Even before the black hole has completely evaporated, the book's information starts to slowly leak out via each piece of Hawking radiation.

Yet a puzzling thing occurs if we think about the same situation from the viewpoint of someone who is falling into the black hole along with the book. In this case, the book seems to simply pass through the boundary of the black hole and stay inside. Thus, to this inside observer, the information in the book is also contained within the black hole forever. On the other hand, we have just argued that from a distant observer's point of view, the information will be *outside*. Which is correct? You might think that the information is simply duplicated: one copy inside and the other outside. Such a solution, however, is impossible. In quantum mechanics, the so-called no-cloning theorem prohibits faithful, full copying of information. Therefore, it seems that the two pictures seen by the two observers cannot both be true.

Physicists Gerard 't Hooft of Utrecht University in the Netherlands, Susskind and their collaborators have proposed the following solution: the two pictures can both be valid but not at the same time. If you are a distant observer, then the information is outside. You need not describe the interior of the black hole, because you can never access it even in principle; in fact, to avoid cloning information, you must think of the interior spacetime as nonexistent. On the other hand, if you are an observer falling into the hole, then the interior is all you have, and it contains the book and its information. This view, however, is possible only at the cost of ignoring the Hawking radiation being emitted from the black hole—but such a conceit is allowed because you yourself have crossed the black hole boundary and accordingly are trapped inside, cut off from the radiation emitted from the boundary. There is no inconsistency in either of these two viewpoints; only if you artificially "patch" the two, which you can never physically do, given that you cannot be both a distant and a falling observer at the same time, does the apparent inconsistency of information cloning occur.

COSMOLOGICAL HORIZONS

THIS BLACK HOLE CONUNDRUM may seem unrelated to the issue of how the many-worlds notion of quantum mechanics and the multiverse can be connected, but it turns out that the boundary of a black hole is similar in important ways to the so-called cosmological horizon—the boundary of the spacetime region within which we can receive signals from deep space. The horizon exists because space is expanding exponentially, and objects farther than this cutoff are receding faster than the speed of light, so any message from them can never reach us. The situation, therefore, is akin to a black hole viewed by a distant observer. Also, as in the case of the black hole, quantum mechanics requires an observer inside the horizon to view spacetime on the other side of the boundary—in this case, the exterior of the cosmological horizon—as nonexistent. If we consider such spacetime in addition to the information that can be retrieved from the horizon later (analogous to Hawking radiation in the black

© 2017 Scientific American

hole case), then we are overcounting the information. This problem implies that any description of the quantum state of the universe should include only the region within (and on) the horizon—in particular, there can be no infinite space in any single, consistent description of the cosmos.

If a quantum state reflects only the region within the horizon, then where is the multiverse, which we thought existed in an eternally inflating infinite space? The answer is that the creation of bubble universes is probabilistic, like any other process in quantum mechanics. Just as a quantum measurement could spawn many different results distinguished by their probability of occurring, inflation could produce many different universes, each with a different probability of coming into being. In other words, the quantum state representing eternally inflating space is a superposition of worlds—or branches—representing different universes, with each of these branches including only the region within its own horizon.

Because each of these universes is finite, we avoid the problem of predictability that was raised by the prospect of an infinitely large space that encompasses all possible outcomes. The multiple universes in this case do not all exist simultaneously in real space—they coexist only in "probability space," that is, as possible outcomes of observations made by people living inside each world. Thus, each universe—each possible outcome—retains a specific probability of coming into being.

This picture unifies the eternally inflating multiverse of cosmology and Everett's many worlds. Cosmic history then unfolds like this: the multiverse starts from some initial state and evolves into a superposition of many bubble universes. As time passes, the states representing each of these bubbles further branch into more superpositions of states representing the various possible outcomes of "experiments" performed within those universes (these need not be scientific experiments—they can be any physical processes). Eventually the state representing the whole multiverse will thus contain an enormous number of branches, each of which represents a possible world that may arise from the initial state. Quantum-mechanical probabilities therefore determine outcomes in cosmology and in microscopic processes. The multiverse and quantum many worlds are really the same thing; they simply refer to the same phenomenon—superposition—occurring at vastly different scales.

In this new picture, our world is only one of all possible worlds that are allowed by the fundamental principles of quantum physics and that exist simultaneously in probability space.

THE REALM BEYOND

TO KNOW IF THIS IDEA is correct, we would want to test it experimentally. But is that feasible? It turns out that discovery of one particular phenomenon would lend support to the new thinking. The multiverse could lead to a small amount of negative spatial curvature in our universe—in other words, objects would travel through space not along straight lines as in a flat cosmos but along curves, even in the absence of gravity. Such curvature could happen because, even though the bubble universes are finite as seen from the perspective of the entire multiverse, observers inside a bubble would perceive their universe to be infinitely large, which would make space seem negatively curved (an example of negative curvature is the surface of a saddle, whereas the surface of a sphere is positively curved). If we were inside one such bubble, space should likewise appear to us to be bent.

Evidence so far indicates that the cosmos is flat, but experiments studying how distant light bends as it travels through the cosmos are likely to improve measures of the curvature of our universe by about two orders of magnitude in the next few decades. If these experiments find any amount of negative curvature, they will support the multiverse concept because, although such curvature is technically possible in a single universe, it is implausible there. Specifically, a discovery supports the quantum multiverse picture described here because it can naturally lead to curvature large enough to be detected, whereas the traditional inflationary picture of the multiverse tends to produce negative curvature many orders of magnitude smaller than we can hope to measure.

Interestingly, the discovery of positive curvature would falsify the multiverse notion presented here because inflation theory suggests that bubble universes could produce only negative curvature. On the other hand, if we are lucky, we may even see dramatic signs of a multiverse—such as a remnant from a "collision" of bubble universes in the sky, which may be formed in a single branch in the quantum multiverse. Scientists are, however, far from certain if we will ever detect such signals.

I and other physicists are also pursuing the quantum multiverse idea further on a theoretical level. We can ask fundamental questions such as: How can we determine the quantum state of the entire multiverse? What is time, and how does it emerge? The quantum multiverse picture does not immediately answer these questions, but it does provide a framework to address them. Lately, for instance, I have found that constraints imposed by the mathematical requirement that our theory must include rigorously defined probabilities may enable us to determine the unique quantum state of the entire multiverse. These constraints also suggest that the overall quantum state stays constant even though a physical observer, who is a part of the multiverse state, will see that new bubbles constantly form. This implies that our sense of the universe changing over time and, indeed, the concept of time itself may be an illusion. Time, according to this notion, is an "emergent concept" that arises from a more fundamental reality and seems to exist only within local branches of the multiverse.

Many of the ideas I have discussed are still quite speculative, but it is thrilling that physicists can talk about such big and deep questions based on theoretical progress. Who knows where these explorations will finally lead us? It seems clear, though, that we live in an exciting era in which our scientific explorations reach beyond what we thought to be the entire physical world—our universe—into a potentially limitless realm. SA

Yasunori Nomura is a professor of physics and director of the Berkeley Center for Theoretical Physics at the University of California, Berkeley. He is also a senior faculty scientist at Lawrence Berkeley National Laboratory and a principal investigator at the University of Tokyo's Kavli Institute for the Physics and Mathematics of the Universe.

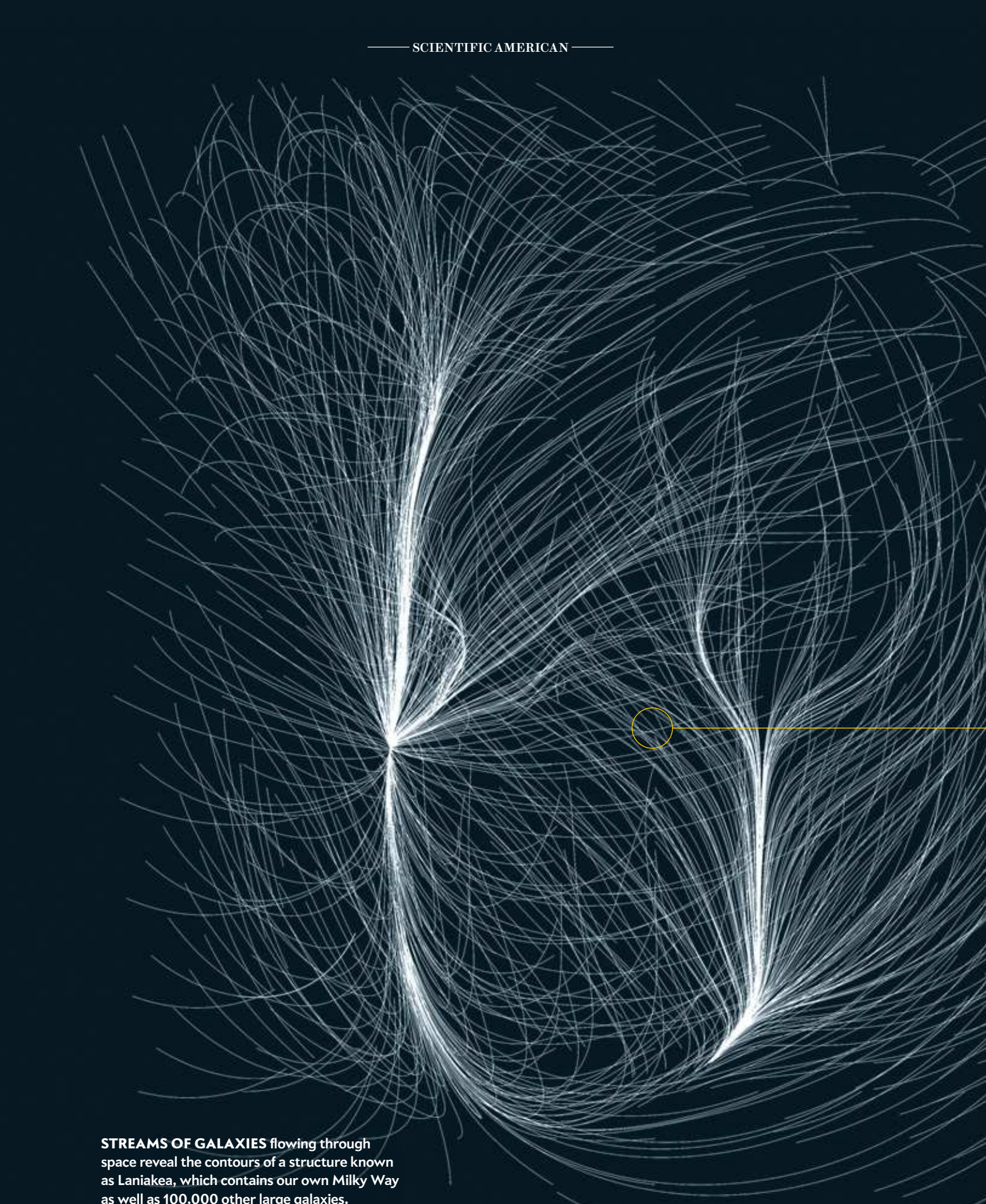
MORE TO EXPLORE

Physical Theories, Eternal Inflation, and the Quantum Universe. Yasunori Nomura in *Journal of High Energy Physics,* Vol. 2011, No. 11, Article No. 063; November 2011. Preprint available at **https://arxiv.org/abs/1104.2324**

Multiverse Interpretation of Quantum Mechanics. Raphael Bousso and Leonard Susskind in *Physical Review D,* Vol. 85, No. 4, Article No. 045007. Published online February 6, 2012. Preprint available at **https://arxiv.org/abs/1105.3796**

scientificamerican.com/magazine/sa

© 2017 Scientific American



STREAMS OF GALAXIES flowing through space reveal the contours of a structure known as Laniakea, which contains our own Milky Way as well as 100,000 other large galaxies.

© 2017 Scientific American

Our Place in the Cosmos

The Milky Way turns out to be part of a massive supercluster of galaxies that forms one of the largest known structures in the universe. This discovery is only the beginning of a new effort to map the cosmos

By
Noam I. Libeskind and R. Brent Tully

© 2017 Scientific American

IMAGINE VISITING A FAR DISTANT GALAXY AND ADDRESSING A POSTCARD TO YOUR loved ones back home. You might begin with your house on your street in your hometown, somewhere on Earth, the third planet from our sun. From there the address could list the sun's location in the Orion Spur, a segment of a spiral arm in the Milky Way's suburbs, followed by the Milky Way's residence in the Local Group, a gathering of more than 50 nearby galaxies spanning some seven million light-years of space. The Local Group in turn exists at the outskirts of the Virgo Cluster, a 50-million-light-year-distant cluster of more than 1,000 galaxies that is itself a small part of the Local Supercluster, a collection of hundreds of galaxy groups sprawled across more than 100 million light-years. Such superclusters are believed to be the biggest components of the universe's largest-scale structures, forming great filaments and sheets of galaxies surrounding voids where scarcely any galaxies exist at all.

Until recently, the Local Supercluster would have marked the end of your cosmic address. Beyond this scale, it was thought, further directions would become meaningless as the boundary between the crisp, supercluster-laced structure of galactic sheets and voids gave way to a homogeneous realm of the universe with no larger discernible features. But in 2014 one of us (Tully) led a team that discovered we are part of a structure so immense that it shattered this view. The Local Supercluster, it turns out, is but one lobe of a much larger supercluster, a collection of 100,000 large galaxies stretching across 400 million light-years. The team that discovered this gargantuan supercluster named it Laniakea—Hawaiian for "immeasurable heaven"—in honor of the early Polynesians who navigated the great expanse of the Pacific Ocean by the stars. The Milky Way sits far from Laniakea's center, in its outermost hinterlands.

Laniakea is more than just a new line on our cosmic address. By studying the architecture and dynamics of this immense structure, we can learn more about the universe's past and future. Charting its constituent galaxies and how they behave can help us better understand how galaxies form and grow while telling us more about the nature of dark matter—the invisible substance that astronomers believe accounts for about 80 percent of the stuff in the universe.

Laniakea could also help demystify dark energy, a powerful force only discovered in 1998 that is somehow driving the accelerating expansion of the universe and thus shaping the cosmos's ultimate fate. And the supercluster may not actually be the final line of our cosmic address—it could, in fact, be part of an even bigger structure yet to be discovered.

PROBING MYSTERIES WITH GALACTIC FLOWS

THE TEAM that discovered Laniakea did not exactly set out to find it. Instead Laniakea emerged from efforts to answer lingering fundamental questions about the nature of the universe.

Scientists have known for nearly a century that the cosmos is expanding, pulling galaxies away from one another like dots moving apart on the surface of an inflating balloon. Yet in recent decades it has become clear that most galaxies are not separating as fast as would be expected if expansion were the only force acting on them. Another, more local force is at work, too—the gravitational tugs from other nearby accumulations of matter, which can offset a galaxy's flow in the expansion of the cosmos. The difference between a galaxy's motion from cosmic expansion and the motion from its local environment is called peculiar velocity.

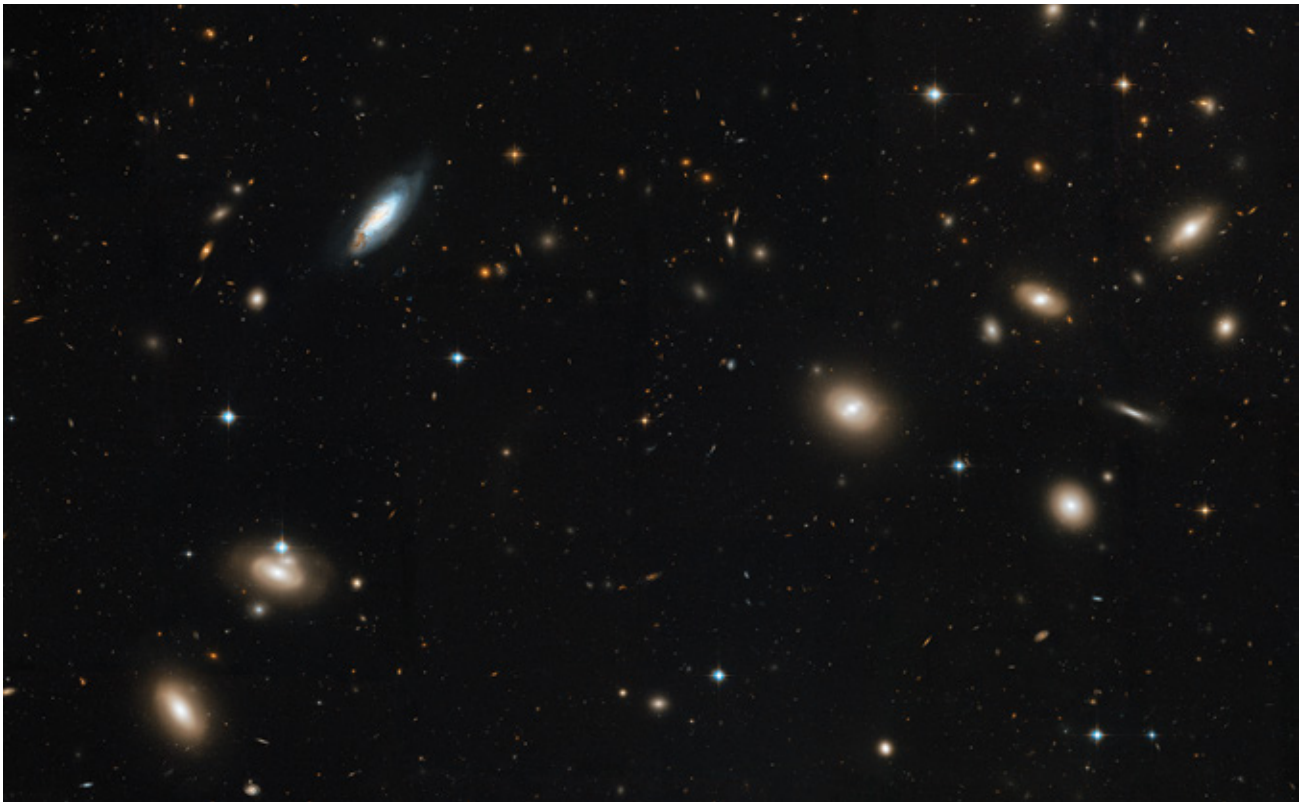
If we take all the stars in all the galaxies that we see and add all the gas and other ordinary matter that we know about, we fall short of explaining the gravitational sources of observed peculiar velocities by an order of magnitude. In our ignorance we astronomers call what is missing "dark matter." We presume this dark matter consists of particles that interact almost exclusively with the rest of the universe through gravity rather than through other forces such as electromagnetism and that the dark matter exerts the "missing" gravitational force needed to account for the observed velocities. Scientists think that galax-

IN BRIEF

Similar to how stars clump together into star clusters and galaxies, galaxies themselves gather into clusters, and galactic clusters group into superclusters.
These galactic superclusters are the building blocks of great filaments, sheets and voids that constitute the largest measurable structures in the universe.
Recent studies of the motions of thousands of nearby galaxies have revealed that the Milky Way's home supercluster is far larger than previously thought. Astronomers call this newfound supersized supercluster "Laniakea."
More detailed mapping of Laniakea and its neighboring superclusters could reveal new details about galaxy formation and help researchers solve the dual cosmological mysteries of dark matter and dark energy.

PRECEDING PAGES: COURTESY OF DANIEL POMARÈDE *CEA Saclay/IRFU*

© 2017 Scientific American



GALAXY CLUSTERS such as this one, the Coma Cluster, are building blocks for the largest structures in the universe. Located more than 300 million light-years away and containing about 1,000 large galaxies, the Coma Cluster is part of an even bigger structure—the Coma Supercluster—which lies beyond Laniakea's boundaries.

COURTESY OF NASA, ESA AND HUBBLE HERITAGE TEAM *STScI/AURA*

ies sit within deep pools of dark matter—dark matter being the invisible scaffolding around which galaxies coalesce.

Tully's group and others realized that creating maps of galactic flows and peculiar velocities could reveal dark matter's hidden cosmic distribution, uncovering the mysterious substance's greatest concentrations by their gravitational influence on galaxies' motions. If, say, streams of galaxies are all flowing toward a particular point, one can assume that the galaxies are being gravitationally drawn to that point by a highly dense region of matter.

They also realized that pinning down the density and distribution of all forms of matter in the universe could help solve another, even deeper mystery: the fact that the cosmos is not only expanding but is doing so at an accelerating rate. This behavior is as counterintuitive as a rock tossed up in the air rushing onward into the heavens rather than falling back to Earth. Whatever powers this bizarre phenomenon is dubbed "dark energy" and has profound implications for the future of the universe. The accelerated expansion suggests the cosmos will ultimately experience a cold death, with most galaxies racing away from one another at accelerating speeds until a final darkness descends on the universe as every star in every galaxy dies and all matter cools to absolute zero. But knowing for sure how it all will end requires not only determining what exactly dark energy is but also how much matter is in the universe: Given a sufficiently high density of matter, in the far future our universe could reverse its expansion to collapse in on itself as a consequence of the self-gravity of its cumulative mass. Or it could instead possess a balanced density of matter that would lead to an infinite but ever slowing expansion.

It was this charting of galactic flows to map the cosmic density of ordinary and dark matter that ultimately led to the unveiling of Laniakea.

FINDING LANIAKEA

MAPPING GALACTIC FLOWS requires knowing both a galaxy's motion stemming from cosmic expansion and its motion stemming from nearby matter. As a first step, astronomers measure a galaxy's redshift—the stretching out of the galaxy's emitted light as it recedes away from us through the expanding universe. A whistle or siren moving toward us has a higher pitch than if moving away because its sound waves are compressed to higher frequencies and shorter wavelengths. Similarly, light waves from a galaxy moving away from us are shifted to lower frequencies and longer, redder wavelengths—the faster they are receding, the more redshifted they become. Thus, a galaxy's redshift gives astronomers a measure of its overall velocity and a rough estimate of its distance.

Astronomers can infer how much of a galaxy's velocity is the result of local gravitational tugging by measuring its distance through other techniques besides redshift. For instance, based on rigorous estimates of the universe's expansion rate, a galaxy measured to be 3.25 million light-years away should have a velocity of about 70 kilometers per second. If instead the galaxy's redshift yields a velocity of 60 kilometers per second, astronomers could infer that matter concentrations near that galaxy are giving it a

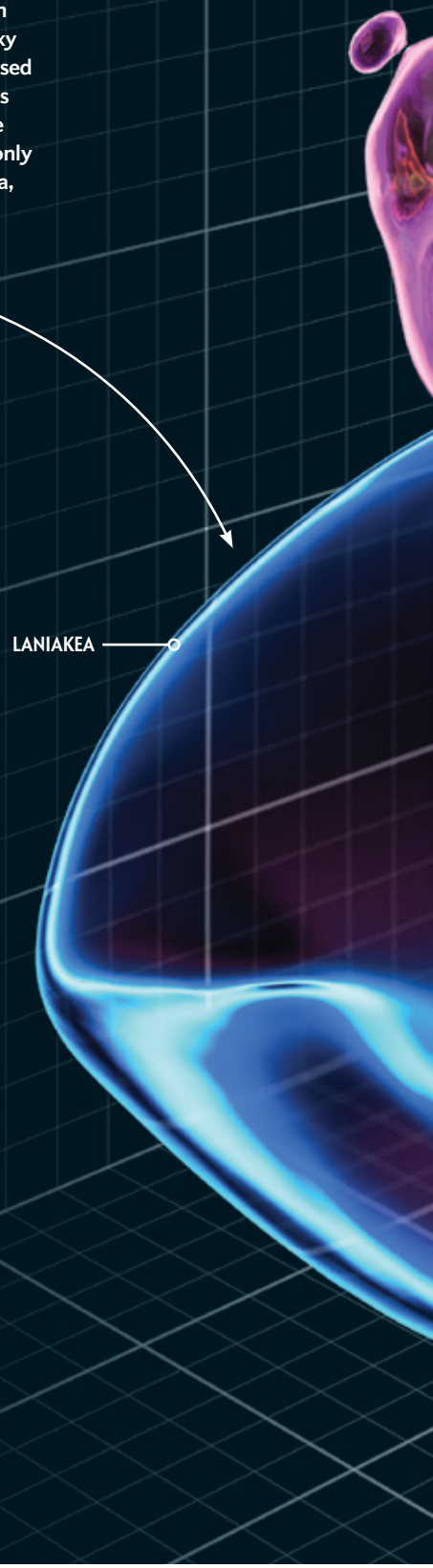
© 2017 Scientific American

Cosmic Landscape

Although they contain hundreds of billions of stars, galaxies are not the largest structures in the universe. Bound together by gravity, hundreds of galaxies can group together as a galaxy cluster. Gravity can concentrate galaxy clusters together, too, forming superclusters composed of hundreds of thousands of galaxies. In this hierarchy, our solar system's cosmic address has traditionally been: the Milky Way galaxy, the Local Group, the Virgo Cluster and, finally, the Local Supercluster. Now, however, new research is revealing that our Local Supercluster is only part of another supercluster more than 100 times larger. This larger supercluster is Laniakea, which in Hawaiian means "immeasurable heaven."

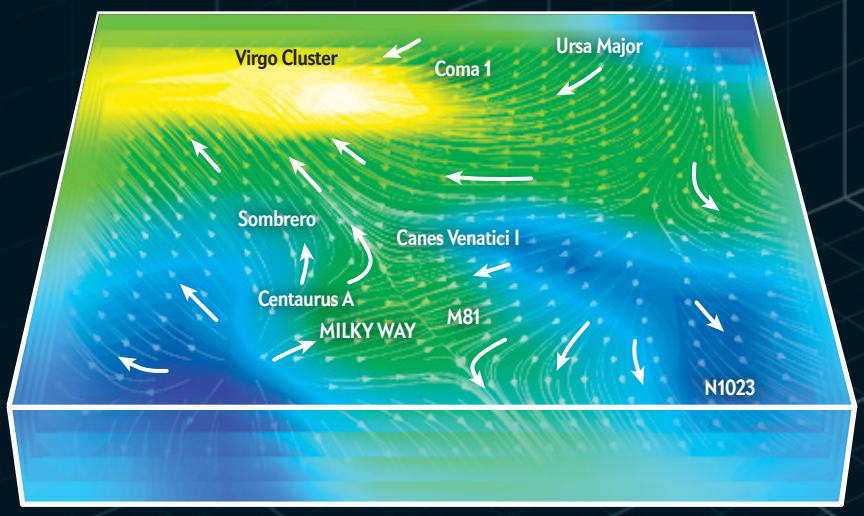
MAPPING THE LANIAKEA SUPERCLUSTER

Considered collectively, the positions and motions of galaxies can either diverge as the universe expands or converge in response to gravity. Supercluster boundaries can be drawn wherever the convergent force of gravity begins to significantly impede the expansion-caused divergent motion of galaxies. Here the positions of more than 8,000 galaxies are mapped and colored according to their relative motions (their velocity and trajectory when both convergent and divergent motion are taken into account). Contours with warmer colors (*yellows* and *pinks*) represent galaxy clusters rapidly converging together. Laniakea's outline appears in cooler blue, delineating where clusters are converging at their slowest. Spanning nearly half a billion light-years, Laniakea encompasses a volume of galaxy clusters that, in the absence of cosmic expansion, would converge to become a single gravitationally bound structure. Beyond Laniakea's boundaries, Shapley, Hercules, Perseus-Pisces and other neighboring superclusters can be seen.



GOING WITH THE FLOW

Zooming in on Laniakea's finer details can yield new insights into the distribution of dark matter and the process of galactic evolution. For instance, consider a three-dimensional slice of Laniakea containing the Milky Way and its enclosing Local Group of galaxies (*detail*). The arrows track the motions of galaxies, which flow like water toward areas where matter is dense and gravity is strong (*depicted in warmer colors*) and away from regions of low density (*cooler colors*). The bulk flows of galaxies reveal cosmic concentrations of matter (be it normal or dark). Flow measurements by Libeskind have shown that the Local Group is falling along a 50-million-light-year-long filament of dark matter toward the Virgo Cluster (*yellow*), a gathering of more than 1,000 galaxies squeezed into a volume of 13 million light-years. Such filaments are thought to play important roles in the formation and evolution of galaxies.



SOURCES: "PLANES OF SATELLITE GALAXIES AND THE COSMIC WEB," BY NOAM I. LIBESKIND ET AL., IN *MONTHLY NOTICES OF THE ROYAL ASTRONOMICAL SOCIETY*, VOL. 452, NO. 1; SEPTEMBER 1, 2015 (*inset info*); DANIEL POMARÈDE, HÉLÈNE M. COURTOIS, YEHUDA HOFFMAN AND BRENT TULLY (*data for Laniakea illustration*)

© 2017 Scientific American

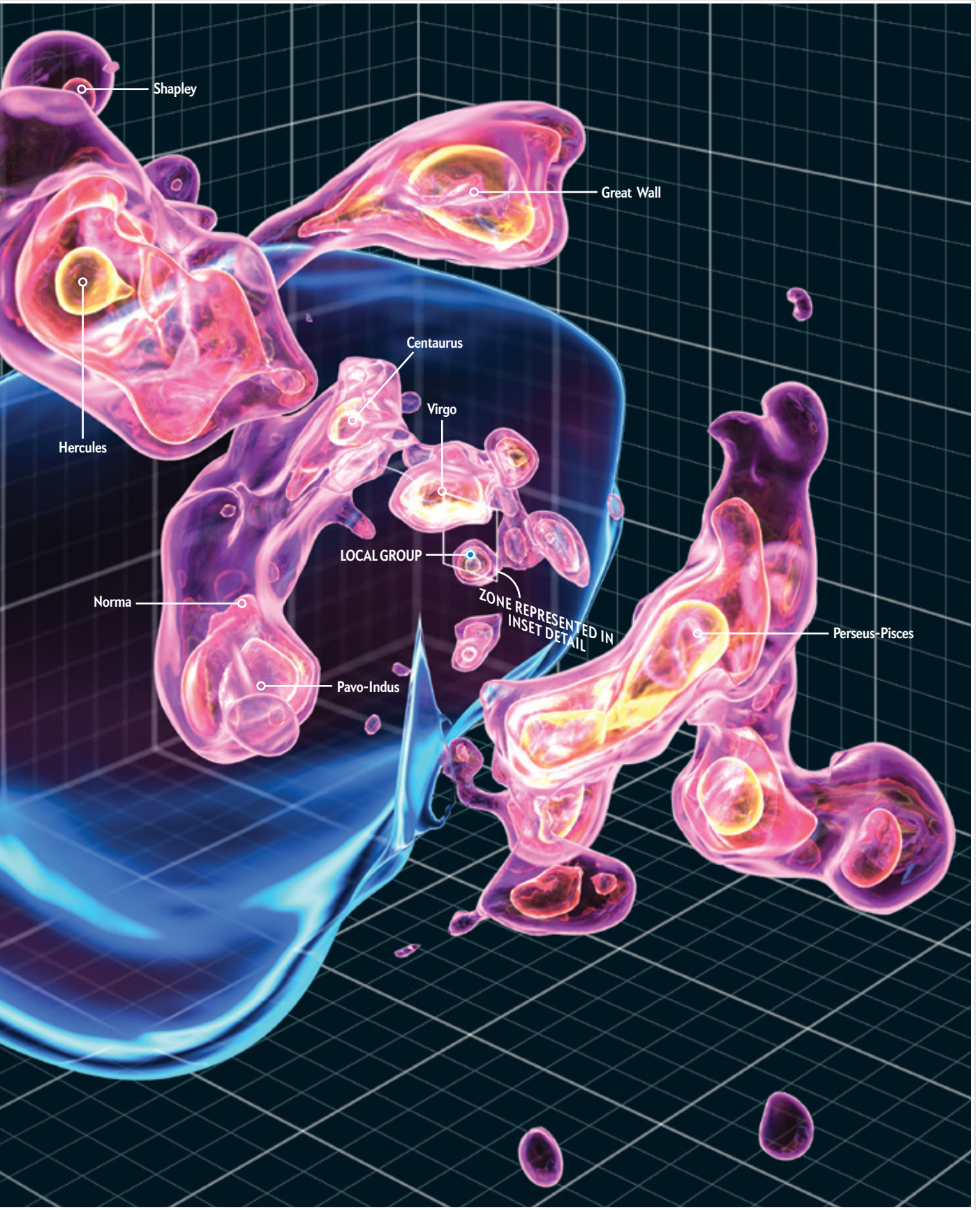


Illustration by Bryan Christie

© 2017 Scientific American

peculiar velocity of 10 kilometers per second. The techniques used to provide redshift-independent distance measurements mostly rely on the fact that light's intensity decreases by the inverse square of the distance from its source. That is, if you see two identical lighthouses but one appears a quarter as bright, then you know the fainter one is two times farther away. In astronomy, such identical lighthouses are called standard candles, astrophysical objects that always shine with the same brightness no matter where they are in the universe. Examples include certain types of exploding or pulsating stars—or even massive galaxies as first proposed by Tully and astronomer J. Richard Fisher in 1977. This Tully-Fisher relation draws on the fact that massive galaxies are both more luminous and rapidly rotating than small galaxies—they have more stars and must spin faster to maintain stability in their stronger gravitational fields. Measure the galaxy's rotation rate, and you learn its intrinsic luminosity; compare that with its apparent luminosity, and you learn its distance.

Each distinct standard candle has a different range where it works best. The pulsating stars called Cepheid variables can only be well observed if the galaxies are close to the Milky Way, and so they are unsuitable for large-scale distance measurements. The Tully-Fisher relation can be used with many spiral galaxies, but the distance estimates they yield have uncertainties of up to 20 percent. Exploding stars called type Ia supernovae yield measurements with half as much uncertainty and shine across vast cosmic distances, but they are rare, only occurring once a century in a good-size galaxy.

If peculiar velocities for a large sample of galaxies in the universe can be obtained, astronomers can then map the largest-scale galactic flows. On these immense scales, the flow of galaxies can be compared with rivers winding through what we call cosmic watersheds, with motions defined by gravitational forces from nearby structures rather than topography. In these "cosmographic" maps, galaxies flow in currents, swirl in eddies and collect in pools to indirectly reveal the structure, dynamics, origins and futures of the largest accumulations of matter in the universe [*see box on two preceding pages*].

Mapping on the scale needed to address our questions about dark matter and dark energy required cataloguing all the best available data from a large number of observational programs. In 2008 Tully, Hélène M. Courtois, now at the Institute of Nuclear Physics of Lyon in France, and their colleagues published the Cosmicflows catalog, which collated multiple data sets to detail the dynamics of 1,800 galaxies within 130 million light-years of the Milky Way. The team expanded its efforts in 2013 with the Cosmicflows-2 catalog, mapping the motions of about 8,000 galaxies within a volume of about 650 million light-years. One member of the team, Yehuda Hoffman of the Hebrew University of Jerusalem, developed methods to precisely derive the distribution of dark matter from the peculiar velocities of the Cosmicflows data.

As the catalog expanded, we were amazed to find an unexpected pattern hidden in the mass of data: the outlines of a new, previously unseen cosmic structure. Clusters of galaxies across a span of more than 400 million light-years all moved together within a local "basin of attraction," akin to water pooling at the lowest point of a landscape's topography. Were it not for the universe's incessant expansion, these galaxies would eventually coalesce into one compact, gravitationally bound structure. All together, this vast swarm of galaxies constituted the Laniakea supercluster.

So far studies of the motions of Laniakea's galaxies show them behaving exactly as would be expected from leading models of dark matter's cosmic distribution—although we cannot see it, we can predict with reasonable accuracy where the universe's invisible stuff accumulates. Furthermore, for better or worse, the total density of visible and dark matter within Laniakea suggests that, just as dark energy theorists thought, the universe is destined for a cold death of ever accelerating expansion.

GALAXIES FLOW IN CURRENTS, SWIRL IN EDDIES AND COLLECT IN POOLS TO REVEAL THE STRUCTURE, DYNAMICS, ORIGINS AND FUTURES OF THE LARGEST ACCUMULATIONS OF MATTER IN THE UNIVERSE.

These conclusions remain provisional. The daunting task of mapping galactic flows still has a long way to go. Currently only 20 percent of galaxies within 400 million light-years also have peculiar velocity determinations, and many standard-candle distance measurements still have large uncertainties. Even so, the emerging map of our galactic neighborhood is giving us a new appreciation of our perch in the cosmographic basins and ranges of the universe.

OUR COSMOGRAPHIC CONTEXT

LET'S TAKE A TOUR of the flowing, rushing components of our newly discovered home, Laniakea, starting with its most familiar part—you. No matter how slow or fast you are traveling on Earth as you read this, you are spinning around the sun, along with the rest of our planet, at about 30 kilometers per second. The sun in turn orbits the galactic center at roughly 200 kilometers per second, and the entire Local Group, including the Milky Way, is hurtling toward a mysterious concentration of mass in the direction of Centaurus at more than 600 kilometers per second (more on this later). You probably never realized you could move so fast simply by reading a magazine article—or doing nothing at all.

As we zoom out from the Milky Way, our journey through Laniakea's expanse begins with two dwarf galaxies, the Small and Large Magellanic Clouds a "mere" 180,000 to 220,000 light-years away. You can glimpse the Magellanics from Earth's Southern Hemisphere, but for the best views you must travel all the way to Antarctica, during the winter. The only other galaxy visible with the naked eye is the giant spiral of Andromeda, although it appears just as a fuzzy patch in a very dark sky.

Andromeda is two and a half million light-years away and is speeding toward us at a peculiar velocity of some 110 kilometers per second. In roughly four billion years it will slam into the Milky Way in a head-on collision and transform both galaxies into a single, featureless ellipsoid of old red stars. It is unlikely that our solar system will be affected during this cosmic car crash—the distance between stars is so large that no two stars are likely to get close enough to collide. The Milky Way, Andromeda and four dozen other galaxies are members of the Local Group, a region where gravity has won the battle against cosmic expan-

© 2017 Scientific American

sion and is undergoing collapse. Like the Milky Way itself, with its Magellanic Clouds, all these large galaxies have their own entourages of dwarfs.

Just beyond the Local Group, within a volume of about 25 million light-years, three distinctive features appear in our maps. Most of the galaxies here, including our own, live in the unimaginatively named Local Sheet. As "sheet" would imply, it is very thin—most of its galaxies are within three million light-years of this structure, itself the equatorial plane of what is referred to as the supergalactic coordinate system. Below this plane, after a gap, is a filament of galaxies—the Leo Spur—as well as galaxies in the so-called Antlia and Doradus Clouds. Above the plane there is mostly nothing nearby. This emptiness is the domain of the Local Void.

If only the galaxies within the Local Sheet are considered, the situation seems very tranquil. These galaxies are flying apart at the rate of the cosmic expansion, with only small peculiar velocities caused by local interactions. Below the Local Sheet, the galaxies of the Antlia and Doradus Clouds and the Leo Spur have small peculiar velocities, too. They are, however, approaching the Local Sheet at high speed. The Local Void is the probable culprit. Voids expand like inflating balloons, and matter moves from underdense to overdense regions to pile up at their boundaries. We now appreciate that the Local Sheet is a wall of the Local Void and that this void is expanding to push us down toward Antlia, Doradus and Leo.

Zooming farther out, we encounter the Virgo Cluster, which is 300 Local Groups' worth of galaxies squeezed into a volume with a diameter of 13 million light-years. These galaxies whiz around at typical speeds of 700 kilometers per second, and any galaxies within 25 million light-years of the cluster's exterior are falling inward to become part of it within 10 billion years. The full extent of Virgo's domain, the region it will eventually capture, extends to a current radius of 35 million light-years. Interestingly, our Milky Way, at 50 million light-years distant, lies just outside this capture zone.

THE GREAT GALACTIC FLOW

THE GREATER REGION around the Virgo Cluster that extends to our location is called the Local Supercluster. Thirty years ago a group of astronomers who became identified by the convivial moniker "the Seven Samurai" discovered that it is not just the Milky Way moving hundreds of kilometers per second in the direction of Centaurus but rather the entire Local Supercluster. They called the mysterious mass pulling all these galaxies together the Great Attractor. In many ways, the Great Attractor is not so mysterious—the density of matter in that direction of the cosmos is obviously high because it contains seven clusters comparable to the Virgo Cluster lying within a sphere 100 million light-years wide. Three of the largest clusters are called Norma, Centaurus and Hydra.

According to our conception of superclusters as cosmic watersheds, which draws their boundaries based on the divergent movements of galaxies, the so-called Local Supercluster is misnamed. It is only part of something bigger—namely, Laniakea, which encompasses other large structures such as the Pavo-Indus filament and the Ophiuchus Cluster. Imagining Laniakea as a city, our traffic-heavy downtown would be the Great Attractor region. As with most urban cores, it is hard to specify a precise center, but an approximation would place it somewhere between the Norma and Centaurus clusters. This positioning puts our Milky Way far out in the suburbs, near the borders of an adjacent supercluster called Perseus-Pisces. This border is so relatively close in cosmic terms that we can study it in detail to define Laniakea's blobby, roughly round, half-billion-light-year-wide boundary. In total, Laniakea's boundaries encompass a mass from both normal and dark matter equivalent to some 100 million billion suns.

Astronomers have been glimpsing the outlines of what may lie beyond Laniakea for decades. Soon after the discovery of the Great Attractor by the Seven Samurai, something even larger emerged from the intergalactic murk. Directly behind the Great Attractor region, but three times farther away, is a monstrous accumulation of clusters—the densest known within the local universe. Because astronomer Harlow Shapley first spied evidence for its existence in the 1930s, this distant, huge structure became known as the Shapley Supercluster. (Incidentally, just like the Local Sheet, the Virgo Cluster and the main band of the Local Supercluster, as well as the Great Attractor and Shapley Supercluster, all lie on the supergalactic equator. Imagine an immense pancake of galactic superclusters, and you have a good picture of our large-scale local environs.)

So what is causing our Local Supercluster's peculiar velocity of 600 kilometers per second? To some degree, the culprit must be the Great Attractor complex. But we must also consider the gravitational pull of the Shapley Supercluster, which is three times farther away but bears four times the number of rich clusters. Now, according to the Cosmicflows-2 compendium—the same catalog that revealed Laniakea—there is even more to the story. The peculiar velocities of the 8,000 galaxies within this catalog demonstrate a coherent flow toward the Shapley Supercluster. This flow encompasses the entire volume of the Cosmicflows-2 catalog, 1.4 billion light-years from end to end. Does it stop there? We do not yet know. Only even bigger surveys mapping even larger swaths of the universe can reveal the ultimate source—and ultimate structure—behind the epic flow of galaxies in our local universe. SA

Noam I. Libeskind is a cosmologist at the Leibniz Institute for Astrophysics Potsdam in Germany. He uses supercomputers to model the evolution of the universe and the formation of galaxies, with a focus on the Milky Way, the Local Group and the small dwarf galaxies that surround us. Follow Libeskind's astrophysics tweets: @satellitegalaxy

R. Brent Tully is an astronomer at the University of Hawaii who, for 40 years, has been measuring the distances to galaxies and mapping their distribution and motions in space. His *Atlas of Nearby Galaxies*, co-written with J. Richard Fisher and published in 1987, is still the most extensive print atlas of the structure of our cosmic neighborhood. Tully takes pride in finding his way without recourse to GPS.

MORE TO EXPLORE

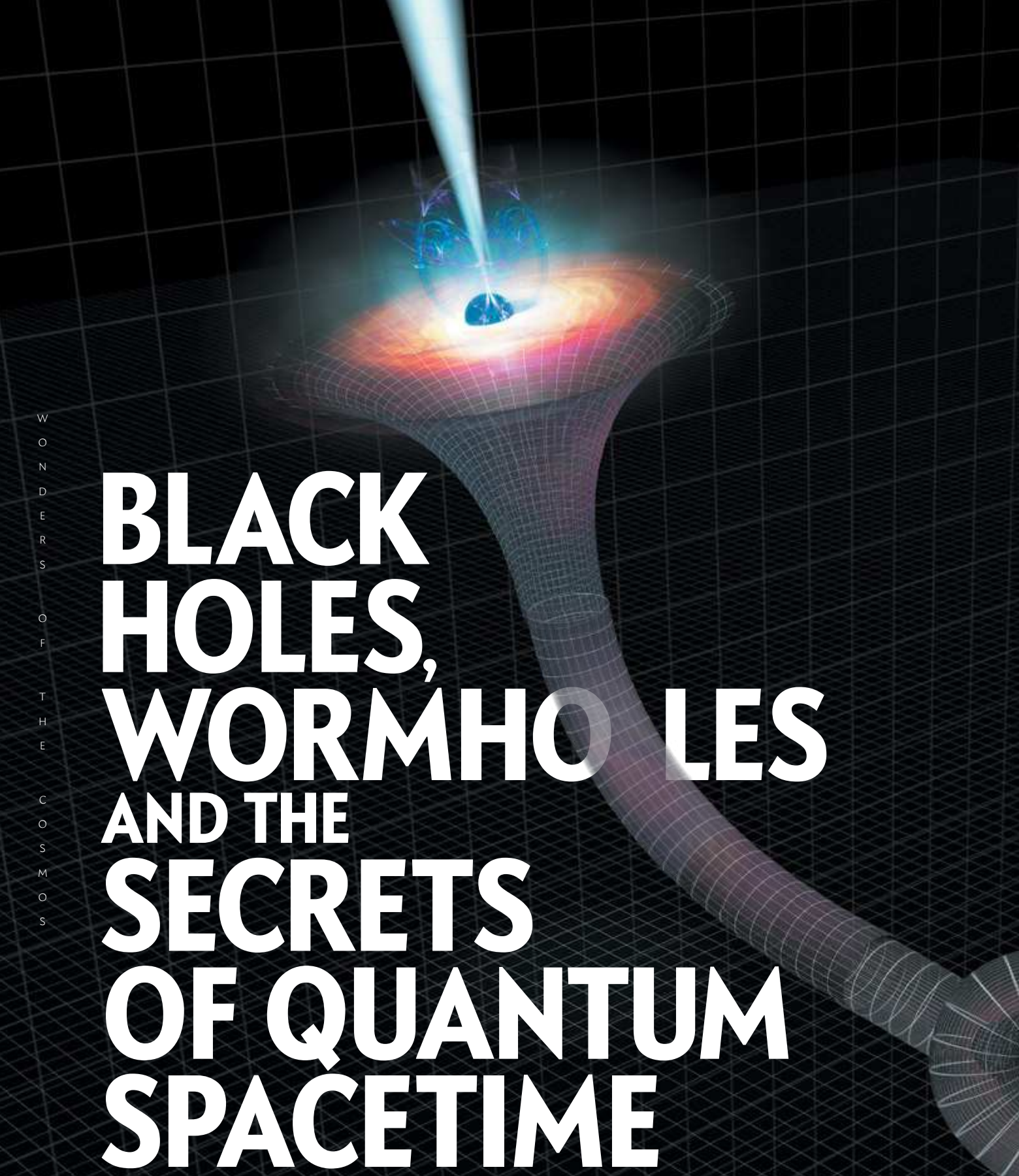
Cosmography of the Local Universe. Hélène M. Courtois et al. in *Astronomical Journal*, Vol. 146, No. 3, Article No. 69; September 2013.

The Laniakea Supercluster of Galaxies. R. Brent Tully et al. in *Nature*, Vol. 513, pages 71–73; September 4, 2014.

Planes of Satellite Galaxies and the Cosmic Web. Noam I. Libeskind et al. in *Monthly Notices of the Royal Astronomical Society*, Vol. 452, No. 1, pages 1052–1059; September 1, 2015.

scientificamerican.com/magazine/sa

© 2017 Scientific American

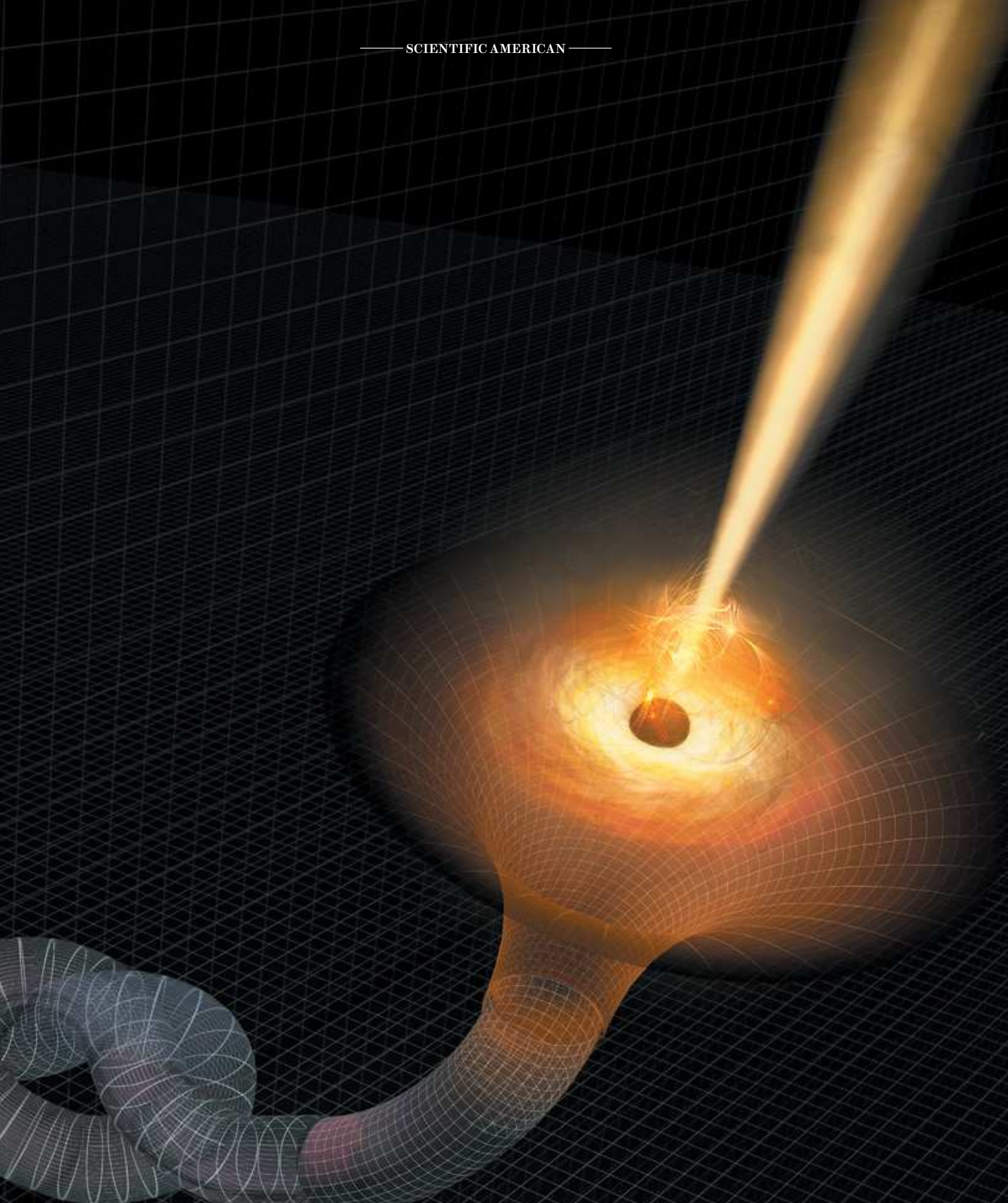


BLACK HOLES, WORMHOLES AND THE SECRETS OF QUANTUM SPACETIME

The weird quantum phenomenon of entanglement could produce shortcuts between distant black holes *By Juan Maldacena*

Illustration by Malcolm Godwin, Moonrunner Design

© 2017 Scientific American



© 2017 Scientific American

THEORETICAL PHYSICS IS FULL OF MIND-BOGGLING IDEAS, BUT TWO OF the weirdest are quantum entanglement and wormholes. The first, predicted by the theory of quantum mechanics, describes a surprising type of correlation between objects (typically atoms or subatomic particles) having no apparent physical link. Wormholes, predicted by the general theory of relativity, are shortcuts that connect distant regions of space and time. Work done in recent years by several theorists, including myself, has suggested a connection between these two seemingly dissimilar concepts. Based on calculations involving black holes, we realized that quantum mechanics' entanglement and general relativity's wormholes may actually be equivalent—the same phenomena described differently—and we believe the likeness applies to situations beyond black holes.

This equivalence could have profound consequences. It suggests that spacetime itself could emerge from the entanglement of more fundamental microscopic constituents of the universe. It also suggests that entangled objects—despite having long been viewed as having no physical connection to one another—may in fact be connected in ways that are far less fantastical than we thought.

Furthermore, the relation between entanglement and wormholes may help in developing a unified theory of quantum mechanics and spacetime—what physicists call a theory of quantum gravity—that derives the physics of the macroscopic cosmos from the laws that govern the interactions of the atomic and subatomic realms. Such a theory is necessary for understanding the big bang and the interior of black holes.

Interestingly, both quantum entanglement and wormholes date back to two articles written by Albert Einstein and his collaborators in 1935. On the surface, the papers seem to deal with very different phenomena, and Einstein probably never suspected that there could be a connection between them. In fact, entanglement was a property of quantum mechanics that greatly bothered the German physicist, who called it "spooky action at a distance." How ironic that it now may offer a bridge to extend his relativity theory to the quantum realm.

BLACK HOLES AND WORMHOLES

TO EXPLAIN WHY I think quantum entanglement and wormholes could be related, we must first describe several properties of black holes, which are intimately involved in this idea. Black holes are regions of curved spacetime that differ drastically from the relatively nondistorted space we are used to. The distinctive feature of a black hole is that we can separate its geometry into two regions: the exterior, where space is curved but objects and messages can still escape, and the interior, lying beyond the point of no return. The interior and exterior are separated by a surface called the event horizon. General relativity tells us that the horizon is just an imaginary surface; an astronaut crossing it would not feel anything special at that location. But having crossed it, a space traveler would be doomed to being squeezed into a region with huge curvature and with no possibility of escape. (In fact, the interior is actually in the future compared with the exterior, so the traveler cannot escape, because he or she cannot travel to the past.)

Just a year after Einstein introduced general relativity, German physicist Karl Schwarzschild found the simplest solution to Einstein's equations describing what would later be called black holes. The geometry that Schwarzschild came up with was so unexpected that it took until the 1960s for scientists to fully understand that it describes a wormhole joining two black holes. From the outside the black holes appear to be separate entities sitting at distant locations, yet they share an interior.

In a 1935 paper, Einstein and his colleague Nathan Rosen, then at the Institute for Advanced Study in Princeton, N.J., anticipated that this shared interior was a kind of wormhole (although they did not understand the full geometry it predicted), and for this reason wormholes are also called Einstein-Rosen (ER) bridges.

The wormhole in Schwarzschild's solution differs from black

IN BRIEF

The laws of quantum physics allow for distant objects to be entangled so that actions on one affect the other, even though they lack a physical link.

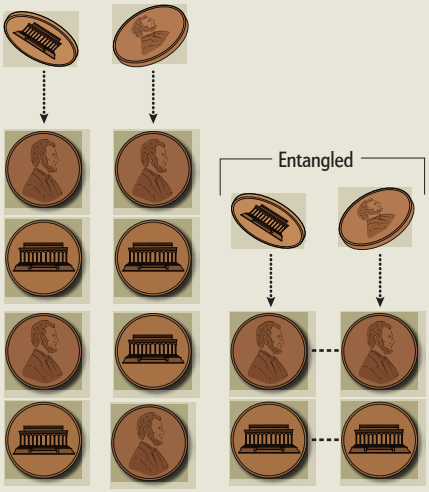
The equations of relativity, which describe the geometry of spacetime, allow for wormholes: shortcuts between distant regions of space and time.

Physicists have suggested that the two phenomena might be equivalent and that this equivalence is a clue for developing a quantum description of spacetime.

© 2017 Scientific American

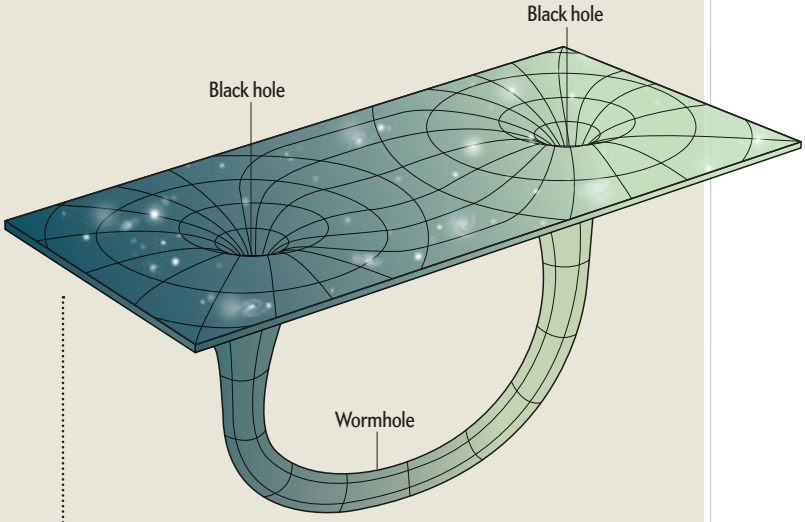
Entanglement Meets Wormholes

Entanglement is a concept from the theory of quantum mechanics describing a special type of correlation between two distant objects. Wormholes, which were predicted by general relativity, are theoretical bridges in spacetime joining distant black holes. Physicists now think that these two phenomena, seemingly disconnected, may be fundamentally related.



ENTANGLEMENT

When two normal coins are thrown, the outcome of one has no effect on the other—any two combinations might result. If two coins are entangled, however, then throwing the first coin determines what will happen to the second. If the first comes out heads, for instance, the second must be heads, and if the first is tails, so must the second be.

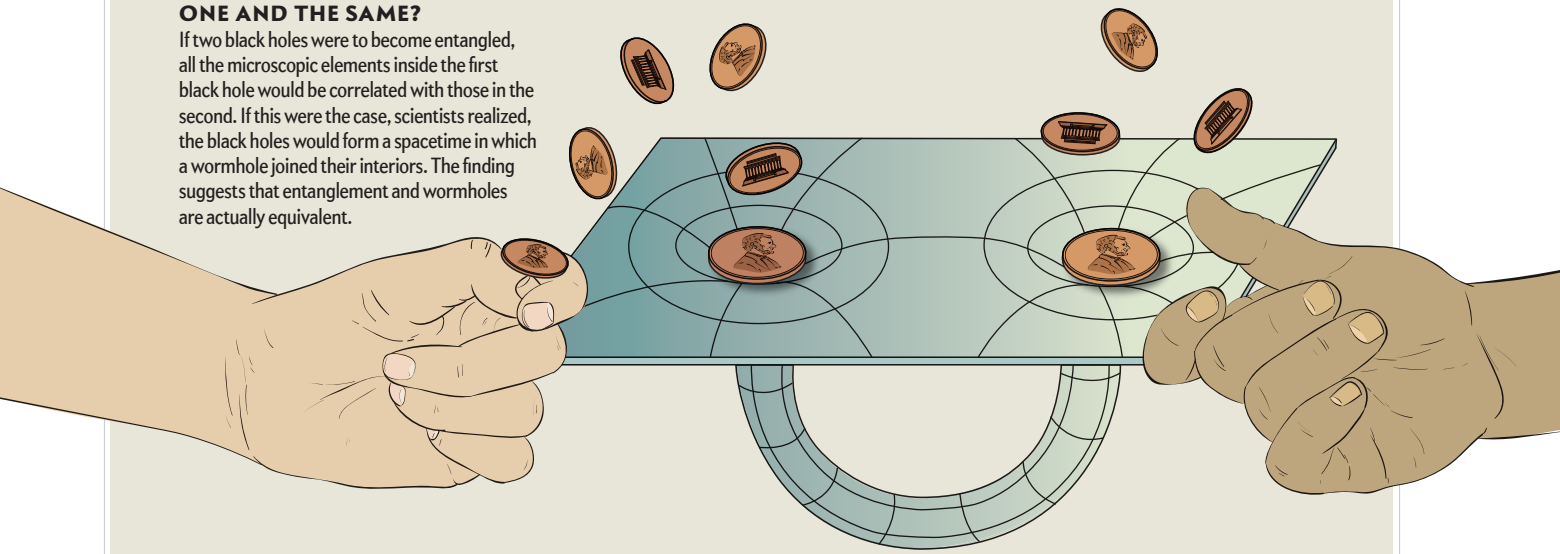


WORMHOLES

The equations of general relativity suggest that wormholes can connect two black holes, even those located vast distances apart, to create a bridge in spacetime. From the outside the two black holes would appear to be separate entities, but they would share an interior connecting them. No person or signal could travel through, however.

ONE AND THE SAME?

If two black holes were to become entangled, all the microscopic elements inside the first black hole would be correlated with those in the second. If this were the case, scientists realized, the black holes would form a spacetime in which a wormhole joined their interiors. The finding suggests that entanglement and wormholes are actually equivalent.



holes that form naturally in the cosmos in that it contains no matter—merely curved spacetime. Because of the presence of matter, naturally formed black holes have only one exterior. Most researchers view the full Schwarzschild solution, with its two exteriors, as a mathematical curiosity irrelevant to the black holes in the universe. Nevertheless, it is an interesting solution and physicists have wondered about its physical interpretation.

The Schwarzschild solution tells us that the wormhole connecting the two black hole exteriors varies with time. It elongates and becomes thinner as time progresses, like stretching out a piece of elastic dough. Meanwhile the two black hole horizons, which at one point touch, separate rapidly. In fact, they pull apart

Illustration by Jillian Ditner

© 2017 Scientific American

so quickly that we cannot use such a wormhole to travel from one exterior to the other. Alternatively, we can say that the bridge collapses before we can cross it. In the dough-stretching analogy, the collapse of the bridge corresponds to the dough becoming infinitesimally thin as it gets stretched more and more.

It is important to note that the wormholes we are discussing are consistent with the laws of general relativity, which do not allow faster than light travel. In that way they differ from science-fiction wormholes that allow instantaneous transport between distant regions of space, as in the 2014 movie *Interstellar*. Sci-fi versions often violate the known laws of physics.

A science-fiction story involving our type of wormhole might look like the following. Imagine two young lovers, Romeo and Juliet. Their families do not like each other and so put Romeo and Juliet on different galaxies, forbidding them to travel. These lovebirds are very smart, however, and manage to construct a wormhole. From the outside the wormhole looks like a pair of black holes, one in Romeo's galaxy and one in Juliet's galaxy. The lovers decide to jump into the interior of their respective black holes. Now, according to their families, they simply committed suicide by jumping in and are never heard from again. Unbeknownst to the outside world, though, the geometry of the wormhole is such that Romeo and Juliet actually meet in the shared interior! And they can live together happily for a while before the bridge collapses, destroying the interior and killing them both.

THE TRADITIONAL WISDOM THAT NOTHING CAN ESCAPE A BLACK HOLE IS TOO SIMPLISTIC.

QUANTUM ENTANGLEMENT

THE 1935 PAPER discussing the other phenomenon of interest to us—entanglement—was written by Einstein, Rosen and Boris Podolsky (also then at the Institute for Advanced Study). The three authors came to be known as EPR. In this famous work, the physicists argued that quantum mechanics allows for the existence of certain strange correlations between distant physical objects, a property that would only later be called entanglement.

Correlations between distant objects can also happen in classical physics. Imagine, for example, that you leave home with a single glove because you forgot the other one at home. Before searching your pocket, you do not know whether you have the left or right glove. Once you see that you have the right-hand glove, though, you will immediately know that the one at home is the lefty. But entanglement involves a different kind of correlation, one that exists between quantities governed by quantum mechanics, which are subject to Heisenberg's uncertainty principle. This principle says that there are pairs of physical variables that are impossible to know accurately at the same time. The best-known example involves the position and velocity of a particle: if we measure its position accurately, its velocity becomes uncertain, and vice versa. EPR wondered what would happen if we decided to measure either the positions or the velocities of the individual particles in a pair separated by a wide distance.

The example that EPR analyzed involves two particles with the same mass moving in a single dimension. Let us call these particles R and J because they are the particles that we will imagine being measured by Romeo and Juliet. We can prepare them in such a way that their center of mass has a well-defined position, which we will call x_{cm}, equal to x_R (the position of R) plus x_J (the position of J). We can require the center of mass to equal zero—in other words, we can say that the two particles are always equidistant from the origin. We can also make the particles' relative velocity, v_{rel}, equal to the velocity of R (v_R) minus the velocity of J (v_J), take a precise value; for example, v_{rel} equals some number we can call v_0. In other words, the difference between the two velocities must stay the same. We are here specifying a position and a velocity accurately but not for the same single object, so we do not violate Heisenberg's uncertainty principle. If we have two different particles, nothing prevents us from knowing the position of the first and the velocity of the second. Similarly, once we fix the position of the center of mass, we cannot say anything about the velocity of the center of mass, but we are free to fix the relative velocity.

Here we get to the most amazing part and the thing that makes quantum entanglement seem so strange. Suppose that our particles are far away from each other, and two distant observers, Romeo and Juliet, decide to measure the particles' positions. Now, because of how the particles have been prepared, if Juliet determines any specific value for x_J, then Romeo will find that his particle's position is the negative of Juliet's ($x_R = -x_J$). Note that Juliet's result is random: the position of her particle will vary from measurement to measurement. Romeo's result, however, is completely fixed by Juliet's. Now suppose they both measure their own particle's velocity. If Juliet gets a specific result for v_J, then Romeo will surely find that his velocity is the value of Juliet's plus the relative velocity ($v_R = v_J + v_0$). Again Romeo's result is completely determined by Juliet's. Of course, Romeo and Juliet are free to choose which variable they will measure. In particular, if Juliet measures the position and Romeo measures the velocity, their results will be random and will not display any correlation.

The strange thing is that even though Romeo's measurements of the position and velocity of his particle are constrained by Heisenberg's uncertainty principle, if Juliet decides to measure the position of her particle, Romeo's particle will have a completely certain position once he knows the result of Juliet's measurement. And the same thing will happen with the velocity. It appears as if, when Juliet measured the position, Romeo's particle immediately "knew" that it must have a well-defined position and an uncertain velocity, whereas the opposite should be the case if Juliet measured the velocity. At first glance this situation appears to allow an instantaneous transmission of information: Juliet can measure the position, and then Romeo would see a definite position for his particle, thus inferring that Juliet measured the position. Romeo would not be able to realize, however, that his particle has a definite position without knowing the actual value of the position that Juliet measured. So in fact, correlations caused by quantum entanglement cannot be used to send signals faster than the speed of light.

Although it has been experimentally confirmed, entanglement may still seem just an esoteric property of quantum systems. Yet during the past two decades these quantum correlations have led to a number of practical applications and breakthroughs in fields such as cryptography and quantum computing.

© 2017 Scientific American

EQUIVALENCE

HOW MIGHT OUR TWO very different, bizarre phenomena—wormholes and entanglement—be related? A further look at black holes points the way to the answer. In 1974 Stephen Hawking showed that quantum effects will cause black holes to emit radiation in the same way a hot object does—proving that the traditional wisdom that nothing can escape a black hole is too simplistic. The fact that black holes radiate implies that they have a temperature—a notion with important ramifications.

Since the 19th century physicists have known that temperature stems from the movement of the microscopic constituents of a system. In a gas, for example, temperature arises from the agitation of molecules. Therefore, if black holes have temperatures, one can expect that they also have some kind of microscopic constituents that collectively are capable of adopting various possible configurations, or so-called microstates. We also believe that, at least as seen from the outside, black holes should behave as quantum systems; that is, they should be subject to all the laws of quantum mechanics. In summary, when we look at a black hole from the exterior we should find a system that can have many microstates, with the probability of its being in any of these configurations essentially equal for each microstate.

Because black holes look like ordinary quantum systems from the outside, nothing prevents us from considering an entangled pair of them. Imagine a couple of very distant black holes. Each has a large number of possible microscopic quantum states. Now imagine an entangled pair of black holes in which each quantum state in the first black hole is correlated with the corresponding quantum state of the second. In particular, if we measure a certain state for the first hole, the other hole must be in exactly the same state.

The interesting thing is that, based on certain considerations inspired by string theory (one approach toward a theory of quantum gravity), we can argue that a pair of black holes with their microstates entangled in this way (that is, in what might be called an EPR entangled state) would produce a spacetime in which a wormhole (an ER bridge) links the interior of both black holes. In other words, quantum entanglement creates a geometric connection between the two black holes. This result is surprising because entanglement, we thought, involves correlations without a physical connection. But the two distant black holes in this case are physically connected through their interior and brought close via the wormhole.

Leonard Susskind of Stanford University and I have called the equivalence of wormholes and entanglement "ER = EPR," because it relates the two articles that Einstein and his colleagues wrote in 1935. From EPR's point of view, the observations near the horizons of each black hole are correlated because the black holes are in a state of quantum entanglement. From ER's vantage point, the observations are correlated because the two systems are linked through the wormhole.

Now, going back to our Romeo and Juliet sci-fi story, we can see what the lovers should do to form an entangled pair of black holes to produce the wormhole. First they need to create many entangled particle pairs, similar to the ones discussed earlier, with Romeo possessing one member of each entangled pair and Juliet the other. They then need to build very complex quantum computers that will manipulate their respective quantum particles and combine them in a controlled way to create a pair of entangled black holes. Such a feat would be terribly hard to achieve in practice, but it seems possible according to the laws of physics. Besides, we did say Romeo and Juliet were very smart!

A UNIVERSAL PRINCIPLE?

THE IDEAS THAT LED US HERE have been developed over the years by many researchers, beginning with a 1976 article by Werner Israel, then at the University of Alberta. There was also interesting work in 2006 on the connection between entanglement and the geometry of spacetime by Shinsei Ryu and Tadashi Takayanagi, both then at the University of California, Santa Barbara. Susskind and I were motivated by research published in 2012 by Ahmed Almheiri, Donald Marolf, Joseph Polchinski and James Sully, all then at U.C. Santa Barbara. They discovered a paradox related to the nature of an entangled black hole's interior. The ER = EPR idea, which says that the interior is part of a wormhole connecting the black hole to another system, alleviates some aspects of this paradox.

Although we identified the connection between wormholes and entangled states using black holes, it is tempting to speculate that the link is more general—that whenever we have entanglement we have a kind of geometric connection. This expectation should hold true even in the simplest case, in which we have only two entangled particles. In such situations, however, the spatial connection could involve tiny quantum structures that would not follow our usual notion of geometry. We still do not know how to describe these microscopic geometries, but the entanglement of these structures might somehow give rise to spacetime itself. It is as if entanglement can be viewed as a thread connecting two systems. When the amount of entanglement becomes larger, we have lots of threads, and these threads could weave together to form the fabric of spacetime. In this picture, Einstein's relativity equations are governing the connections and reconnections of these threads; quantum mechanics is not just an add-on to gravity—it is the essence of the construction of spacetime.

For now, this picture is still wild speculation, but several clues point toward it, and many of us physicists are pursuing its implications. We believe that the seemingly unrelated phenomena of entanglement and wormholes might in fact be equivalent and that this equivalence provides an important clue for developing a description of quantum spacetime—and a long-awaited unification of general relativity and quantum mechanics. SA

Juan Maldacena is a theoretical physicist at the Institute for Advanced Study in Princeton, N.J. He is known for his contributions to the study of quantum gravity and string theory. In 2012 he received a Breakthrough Prize in Fundamental Physics.

MORE TO EXPLORE

Can Quantum-Mechanical Description of Physical Reality Be Considered Complete? A. Einstein, B. Podolsky and N. Rosen in *Physical Review,* Vol. 47, No. 10, pages 777–780; May 15, 1935. http://journals.aps.org/pr/pdf/10.1103/PhysRev.47.777

The Particle Problem in the General Theory of Relativity. A. Einstein and N. Rosen in *Physical Review,* Vol. 48, No. 1, pages 73–77; July 1, 1935. http://journals.aps.org/pr/pdf/10.1103/PhysRev.48.73

Cool Horizons for Entangled Black Holes. J. Maldacena and L. Susskind in *Fortschritte der Physik,* Vol. 61, No. 9, pages 781–811; September 2013. Preprint available at http://arxiv.org/abs/1306.0533

scientificamerican.com/magazine/sa

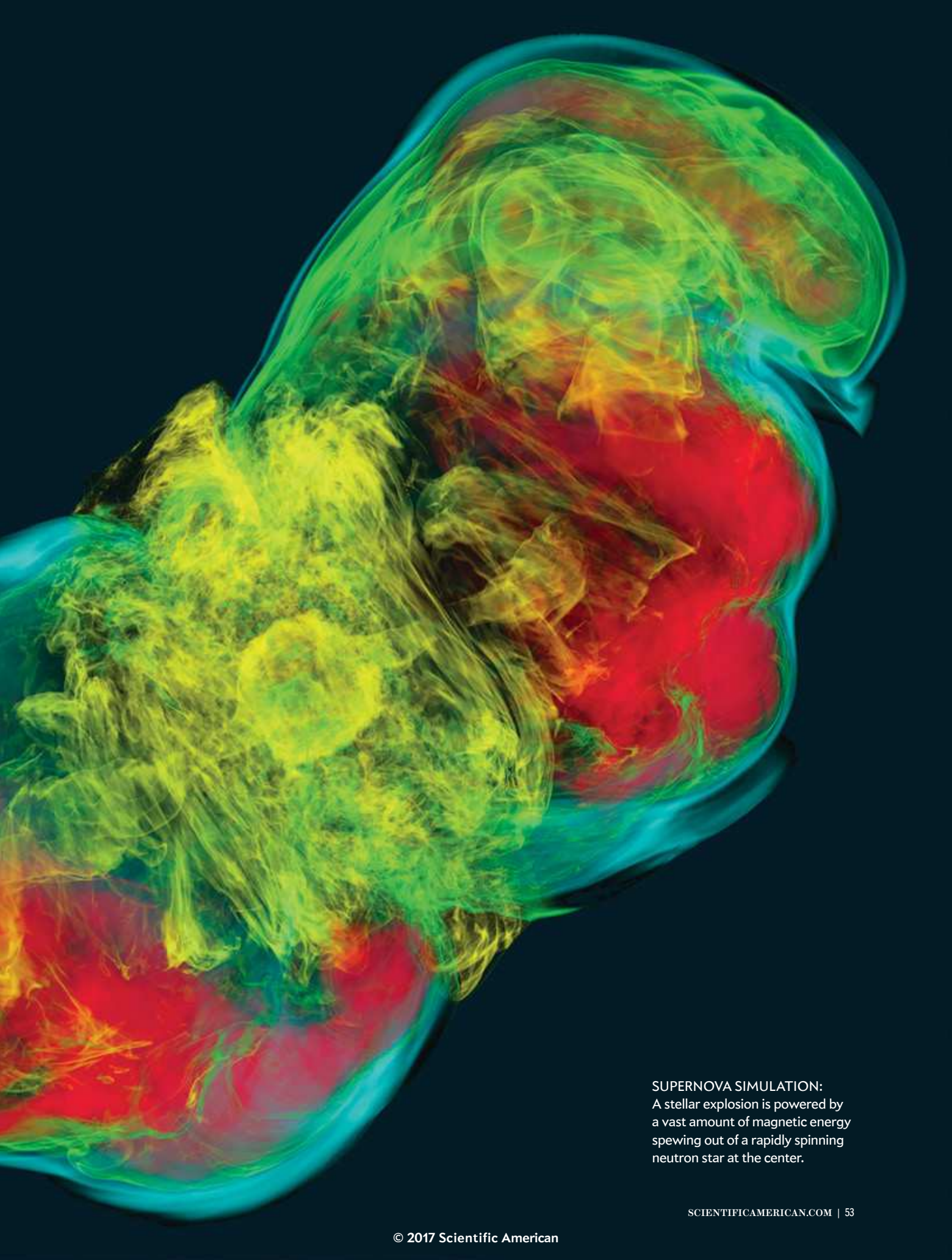
© 2017 Scientific American

STELLAR FIREWORKS

Every year thousands of exploding stars appear in a bizarre assortment of forms. Astronomers want to know what makes them go boom

By Daniel Kasen

© 2017 Scientific American



SUPERNOVA SIMULATION:
A stellar explosion is powered by a vast amount of magnetic energy spewing out of a rapidly spinning neutron star at the center.

© 2017 Scientific American

ROUGHLY EVERY SECOND, SOMEWHERE IN OUR OBSERVABLE UNIVERSE, ANOTHER sun is destroyed in a stellar catastrophe—when a star pulsates, collides, collapses to a black hole or explodes as a supernova. This dynamic side of the universe, lost in the apparent calm of the night sky, has lately come to the forefront of astronomical research. For almost a century scientists have tried to trace what has happened over billions of years of cosmic evolution, but it is only recently that we have begun to parse celestial events on timescales of days and hours and so witness the volatile lives and deaths of stars.

Although in the past we have lacked tools to study these phenomena in detail, evidence of the universe's transience has been around for millennia, going back at least to Chinese observations in A.D. 1006 of a "guest star" that became visible to the naked eye for a few weeks before fading away. The great astronomer Tycho Brahe recorded a similar event in 1572, as did Johannes Kepler about 30 years later. We now understand these apparitions to be the supernova explosions of stars. At their peak, supernovae can shine brighter than a billion suns, but because most occur very far away, they appear to us as dim specks of light, easily lost in a big sky.

Modern technology is now revolutionizing the study of the dynamic universe. Telescopes have become robotic and outfitted with high-resolution digital cameras that feed data to computerized image processing and pattern-recognition software. The machines monitor large swaths of sky on a regular basis, keeping a digital eye on anything that goes bump in the night. Over the past decade or so this newfound technological capability has enabled astronomers to discover thousands of new stellar explosions every year—each week we find as many new supernovae as had been seen in the entire 20th century.

Not only are we collecting more supernovae—we are uncovering bizarre new species. Some stellar explosions shine exceedingly bright, 100 times more luminous than ordinary supernovae; others are 100 times as dim. Some are colored deep red; others are ultraviolet. Some shine brightly for years; others fade away in a few days. Stellar deaths are turning out to be vastly more diverse than we had realized.

Astronomers are still trying to figure out what drives these odd stellar explosions. Clearly, they are telling us something important about the lives and deaths of stars and about physics under the most extreme conditions of temperature, density and gravity. By studying the full menagerie of supernovae, we hope to finally learn what causes stars to crumble and transform into dead stellar remains such as black holes.

Supernovae also have something to teach us about our own origins. After the big bang the universe contained mostly the lightest of atoms: hydrogen and helium. According to theory, everything else we encounter—the calcium in our bones, the iron in our blood—was fused and expelled in exploding stars. Scientists used to think that run-of-the-mill supernovae created all of the heaviest elements, but the discovery of so many off-kilter explosions now suggests that different squares on the periodic table may have different points of origin. By observing large numbers of diverse supernovae, we are getting closer to pinning down how an assortment of stellar explosions may have contributed to the blend of elements that make up our planet and all its life.

STELLAR CATASTROPHE

TO UNDERSTAND HOW odd some of the supernovae we are discovering are, let us first consider the typical supernova, which is itself a truly remarkable phenomenon. A star is a type of stable nuclear reactor: a massive ball of plasma, bound together by gravity and powered by nuclear fusion in its compressed core. The heat from fusion provides a pressure that counteracts the inward pull of gravity. A supernova explosion represents some kind of catastrophic instability in this balance of forces—the runaway victory of gravity over nuclear burning, or vice versa.

The most common type of supernova occurs in moderately sized stars containing 10 or more times the mass of our sun. These stars live for millions of years, continually fusing hydrogen into progressively heavier elements. Once they have burned their insides to iron, which is essentially nuclear ash, fusion cannot continue. Without this outward pressure, the innermost core of the star collapses under the pull of gravity and compresses a mil-

IN BRIEF

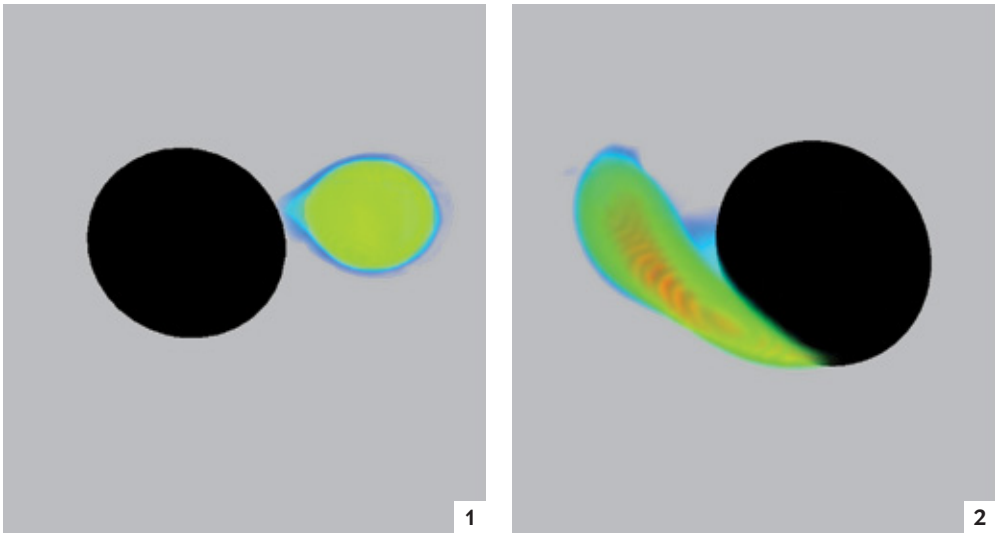
The advent of automated telescopes that quickly scan the sky has revealed a wealth of unusual supernovae, including stellar explosions that are 100 times brighter than normal, as well as unexpectedly dim supernovae that barely make a bang.

Scientists have several theories about what types of stars might be creating these oddball events and how they might give rise to the heavy chemical elements that populate our planet and our bodies.

Future observations should help answer many fundamental questions about how stars live and die and how they affect life on Earth.

PRECEDING PAGES: COURTESY OF PHILIPP MOESTA *University of California, Berkeley*

© 2017 Scientific American



COLLISION: A neutron star (*green*) merges with a black hole in this simulation. The process is similar to what scientists think happens when two neutron stars coalesce—a scenario that could explain some underluminous supernovae. Here the neutron star gets distorted by the black hole's gravity (*1* and *2*). A bit of matter is ejected in a tail (*3*), and the rest wraps around the black hole (*4*). Ultimately almost all of the neutron star is swallowed.

COURTESY OF FRANCOIS FOUCART *Lawrence Berkeley National Laboratory and University of California, Berkeley*

lionfold in volume, transforming into an ultradense nugget called a neutron star, which packs a mass greater than that of the sun into a region no more than a few miles across. The enormous energy released in the free fall blows apart the rest of the star.

To get a sense of the energy involved in a typical supernova explosion, imagine that our sun burned its entire supply of hydrogen—enough fuel to sustain it for more than 10 billion years—within a few seconds. That enormous amount of energy is quantified by its own physical unit: one bethe (named after Nobel Prize winner Hans Bethe). When a supernova explodes, the star's interior temperature rises above five billion degrees Fahrenheit, driving a supersonic blast wave that leaves in its wake a mess of freshly fused heavy elements, such as silicon, calcium, iron, and radioactive isotopes of nickel, cobalt and titanium. In a matter of minutes, the star blows apart into a cloud of ash and radioactive debris, expelled at 20 million miles per hour, or a few percent the speed of light.

Our own sun is, fortunately, too small to ever go supernova, but if it did, the first indication on Earth would be a brief burst of intense x-rays that would kill all life on the planet. Within a few minutes the solar debris cloud would double in size and become about 1,000 times brighter than the sun. After a few hours the cloud would engulf Earth, and a day later it would swallow Jupiter and Saturn. After a few weeks the solar ashes would spread across the entire solar system. By this time, the debris cloud would finally become translucent, and the bottled-up light would flood out, reaching a peak brightness of about one billion solar luminosities before fading away.

Astronomers almost never catch the brief x-ray blast of a supernova explosion itself, and only rarely can we dig up an archived picture of the original star before it exploded. What we typically see is only the aftermath: that giant cloud of expanding, radioactive debris that glows visibly for weeks or longer. By examining the ashes, we try to piece together a story of what type of star was destroyed and why.

STRANGELY BRIGHT

AMONG THE ZOO of weird supernovae recently discovered, perhaps the most dramatic are hyperenergetic explosions—what I will call ultranovae—that are more than 100 times as luminous

© 2017 Scientific American

as ordinary supernovae. They are the brightest and most distant supernovae ever discovered, visible most of the way across the observable universe. Such events are extremely rare; perhaps one goes off for every 1,000 ordinary supernovae. Astronomers have no conclusive evidence to explain why these blasts are so bright, but there are three leading theories. One of them may explain most or all of the ultranovae we see, but it is more likely that all three scenarios occur with some frequency.

PARTICLE-PAIR SUPERNOVAE. Naturally, many try to associate ultranovae with extremely massive stars. Theory suggests that very large stars are actually rather delicate characters, susceptible to a variety of instabilities. In particular, stars between 150 and 250 solar masses may become so hot in their cores that they generate a flurry of matter-antimatter particle pairs (namely, electrons and positrons). Producing these particles costs energy, which saps a star's outward pressure and causes its core, still loaded with combustible nuclear fuel, to fall in. The result would be disastrous. The compression of the core would accelerate nuclear fusion out of control, burning almost everything in sight. The sudden energy release—as much as 100 bethes' worth—would reverse the collapse and explode the star completely. Nothing would be left behind.

These most giant of nuclear explosions would produce debris clouds 1,000 times more radioactive than those of ordinary supernovae. Because the clouds are also thought to be extremely massive and opaque, the light would take a year or longer to diffuse out. We therefore expect the aftermath of these explosions to be extremely bright and long-lasting. A few of the recently discovered ultranovae have these very properties, leading some astronomers to claim that we have witnessed a giant star killed by an infestation of microscopic particle pairs. Others disagree, arguing that the data are better explained with different theories. Future observations of such bright and long-lived events will, it is hoped, better reveal the composition and speed of the stellar debris cloud and tell us whether this scenario truly does take place.

FALSE ALARM SUPERNOVAE. An alternative idea to explain ultranovae is that they originate in stars slightly lower in mass (around 70 to 150 solar masses). These stars are thought to be subject to similar instabilities as their more massive brethren, but the conditions are often not as severe; after the star begins to crumple and ignite excess burning, it may rebound, reexpand and halt the nuclear reactions before they run out of control, recovering to live another day. In the process of regaining its balance, however, the star will likely blow off a good chunk of its outer layers, producing a supernova "imposter"—an outburst that resembles a dim supernova but in reality is just a near-death experience.

Stars that are in this mass range may go through several of these hiccups, each time losing a bit more matter, until they finally exhaust their nuclear fuel and explode like ordinary supernovae. When such a star finally does die, it will expel debris into an environment that is polluted with shells of material from previous outbursts. The violent collision of the supernova debris cloud with these shells should produce extremely bright fireworks that could explain some of the ultranovae.

Automated surveys have recently recorded the manic last years in the life of a massive star. In 2009 astronomers noted what appeared to be a fairly ordinary, if dim, supernova. Named SN2009ip, it faded away in a few weeks and was largely forgotten. One year later, to everyone's surprise, another dim "supernova" was observed in the exact same location. Apparently the star was not dead yet. In 2012 astronomers saw a third outburst and then, a month later, a very luminous one.

Some scientists believe that the penultimate burst was the true death of the star while the final, most luminous flare resulted from the supernova debris cloud ramming into material from the previous near-death gasps. Others think that the star is still alive and will continue to entertain us with further outbursts. It will take a few years for the dust to clear, but for now we have seen the kind of violent instability we think makes up the end of life for some massive stars.

MAGNETIC SUPERNOVAE. Finally, an alternative line of thinking on ultranovae argues that their excessive brightness has less to do with extreme mass and more to do with extreme rotation. Stars that start off with more than 10 solar masses most likely produce ordinary supernovae that form neutron stars when they die. If such a star was initially rotating rapidly, the collapse might spin up the neutron star to extreme speeds, like a twirling figure skater who brings his or her arms in to accelerate. In principle, a neutron star can spin as rapidly as 1,000 rotations a second—much faster, and the star would be torn apart by centrifugal forces. The kinetic energy stored in such a massive, spinning top is enormous—up to 10 bethes.

How can this spin energy be tapped to power an ultranova? Neutron stars have immense magnetic fields that may transport the energy. To understand how, imagine spinning a refrigerator magnet in your palm. As you do so, you twist up the magnetic field surrounding it. Although you cannot see or feel it, a bit of your expended energy is carried off into space in the form of electromagnetic ripples. We think the same process occurs, on a much larger scale, around neutron stars. The most visually captivating example is the Crab nebula, the remains of a supernova reported by Chinese astronomers in A.D. 1054. Today the light we see from the nebula is powered by a spinning neutron star that generates a whirlpool of magnetized plasma. Over a period of 1,000 years the twisted-up magnetic field has extracted the neutron star spin energy and heated the surrounding gas, powering the beautiful display.

About five years ago my colleague Lars Bildsten of the University of California, Santa Barbara, and I suggested that a souped-up version of this process may explain the high luminosity of ultranovae. The neutron star would need to host magnetic fields 100 to 1,000 times stronger than the one in the Crab nebula and spin near the breakup speed limit. For such a star, nearly the entire spin energy could drain within a month, causing the supernova debris cloud to shine a million times more brightly than the Crab nebula. Although the numbers sound extreme, we have already observed some neutron stars with comparable magnetic fields (though none yet in the supernova stage). They are called magnetars, and they harbor the strongest known magnetic fields in the universe. Ultranovae may therefore sometimes signal the birth and prompt spinning down of a rapidly rotating magnetar.

UNEXPECTEDLY DIM

ON THE OPPOSITE END of the spectrum from ultranovae, astronomers have also recently discovered the strange phenom-

© 2017 Scientific American

Supernova Zoo

Supernova explosions, which signal the deaths of stars, come in a much wider variety than scientists thought. Recent observations have revealed supernovae that are 100 times brighter than usual, as well as underperforming blasts that are 1/100th as bright as the norm. Theorists have several ideas about what types of stars and situations give rise to some of these unusual eruptions.

Ordinary
In a typical supernova, the core of a star weighing 10 or more times the mass of our sun collapses into a dense remnant called a neutron star, and the outer layers explode in a supersonic blast wave.

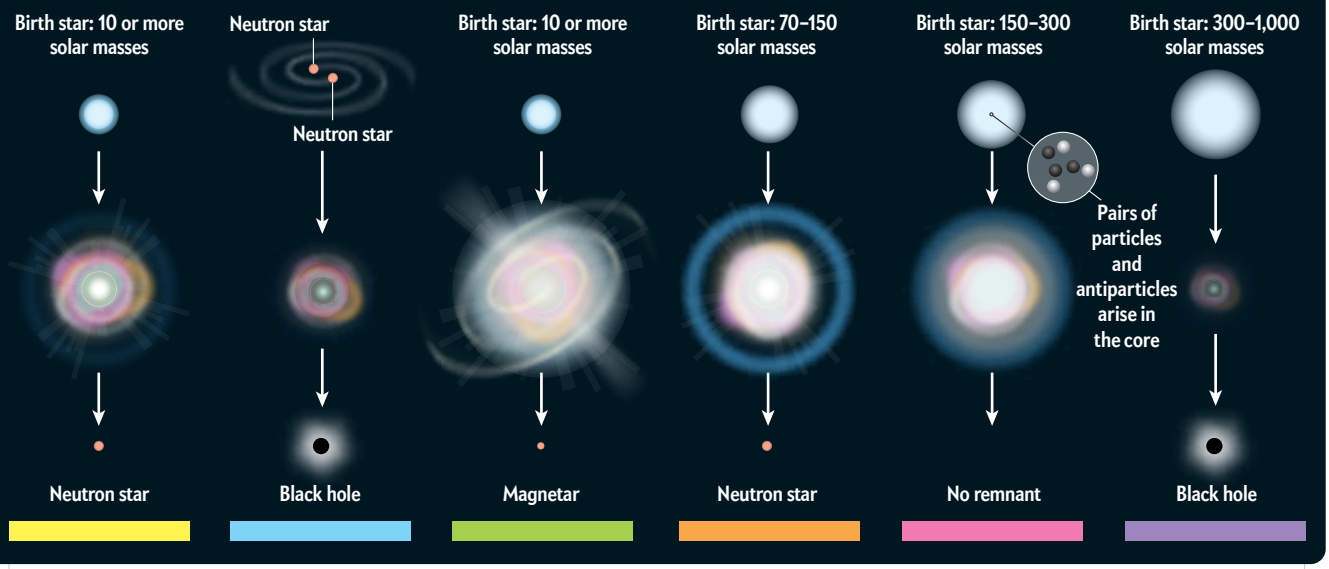
Neutron Star Collision
When two neutron stars crash together, scientists think most of their mass will create a black hole, but a small part may escape to power an under-bright "kilonova."

Magnetar
The collapse of a rapidly rotating star may produce a highly magnetized neutron star, called a magnetar, that taps the spin energy to power an ultraluminous supernova.

False Alarm
A star may begin to go supernova but regain its balance, blowing off just some of its outer layers. When this star finally does explode, the debris will hit the previously shed layers, creating an ultrabright flash.

Particle Pair
The hot cores of very massive stars may give rise to pairs of matter and antimatter particles that precipitate a premature blast. The energy would destroy the star and prevent a black hole from forming.

Complete Collapse
The most massive stars might produce a whimper rather than a bang because their extreme gravity would pull nearly their whole mass into something denser than a neutron star: a black hole.



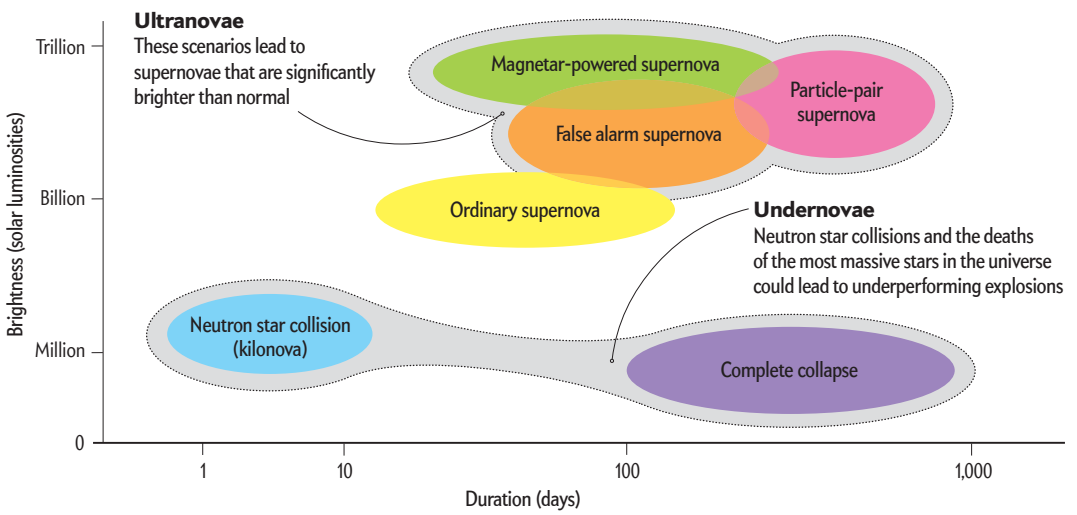


Illustration by Jen Christiansen

© 2017 Scientific American

CRAB NEBULA: This supernova remnant is powered by a spinning neutron star that injects a whirlpool of magnetized plasma (*visible in blue*).

COURTESY OF NASA/CXC/SAO/F. SEWARD (*composite*); COURTESY OF NASA/ESA/ASU/J. HESTER AND A. LOLL (*optical*); NASA/JPL-CALTECH/UNIVERSITY OF MINNESOTA/R. GEHRZ (*infrared*)

enon of underperforming supernovae. Wide-field surveys have found peculiar supernovae that are 100 times as dim as ordinary events. Scientists debate what causes these weak outbursts but suspect some are, surprisingly, the muffled last gasps of the most massive stars to have ever lived.

FAILED SUPERNOVAE. It is unclear just how massive a star can get, but conceivably some may be as large as 300 to 1,000 times the mass of the sun (even larger than the very massive stars we think might explode because of particle pairs). One might expect these mammoths to produce the most spectacular supernova explosions of all. Actually they are probably duds. The gravity of such a star is so strong that once it becomes unstable, total collapse is inevitable. The infall should eventually tear a hole in spacetime itself, forming something denser than a neutron star: a black hole.

Theoretical models show that the bulk of the star will be swallowed by the black hole and suddenly disappear from sight. This hypothetical nonevent has been called an unnova. The models suggest that it is not only extremely massive stars that become supernova flops—some moderately sized stars also fail to explode and shrivel into black holes. Automated surveys are looking for

© 2017 Scientific American

unnovae in a backward way: searching not for a sudden light in the sky but for a bright star that goes lights out in an instant.

Although they fail to make a bang, some of these black hole–forming stars may at least manage a whisper. The cores of certain giant stars are surrounded by a loose, puffy halo of hydrogen gas. As the bulk of such a star gets sucked below the black hole horizon, the halo of gas may heat up and blow away, leading to a faint glow. The death of a very big star would then produce, ironically, a remarkably weak and dim supernova.

COLLIDING NEUTRON STARS. Another type of underluminous eruption might come from a very different kind of extreme event: the collision of two neutron stars. Massive stars are frequently born as orbiting pairs. These stars will go supernova one after the other, and if the pair is not flung apart, what remains is a binary system of two neutron stars (or a neutron star and a black hole or two black holes). Over time, the two compact objects should spiral in closer and closer, ultimately colliding and coalescing into a larger black hole. This process was recently confirmed by the discovery of gravitational waves produced in the merger of two black holes. When neutron stars meld, calculations suggest that the extreme gravitational forces (about 10 billion times the pull of Earth's gravity on our bodies) are strong enough to strip off about 1 percent of the stars' skin and fling it out into space (the remaining 99 percent goes into the black hole).

THE DEATH OF A VERY BIG STAR WOULD PRODUCE A REMARKABLY DIM SUPERNOVA.

This small bit of material that escapes the black hole most likely is peculiar stuff—a vaporized sea of disassociated particles, mostly neutrons, along with some protons and electrons. As the gas decompresses, the particles should begin to bind together into heavier nuclei. The protons will repulse one another because of their positive electric charge, but the neutrons have no charge and will attach to other particles more easily. By progressive addition of neutrons, the nuclei should grow heavier and heavier, producing a shower of elements across the lower half of the periodic table, such as gold, platinum and mercury, mixed into a pool of assorted radioactive waste, including uranium and thorium. Neutron star collisions are one of the few places in the universe scientists think that these heavy elements can be formed.

The abundance of radioactive material should cause the debris cloud to glow like a supernova. But because of the relatively small mass involved (less than 1 percent of that found in a supernova), we expect the light to be about 100 times as dim as an ordinary supernova and to last for only a few days. Recent theoretical work I did with my graduate student Jennifer Barnes at the University of California, Berkeley, suggests that the peculiar composition of heavy metals in such clouds should give the glow a distinctive color, either a deep-red or infrared hue. The phenomenon has been called a kilonova.

Lately astronomers may have, for the first time, seen this radioactive red "smoke" from a neutron star collision. In June 2013 a brief burst of gamma rays alerted astronomers to a possible nearby neutron star merger. They pointed the Hubble Space Telescope at the site and caught a brief infrared glow. A few weeks later it was gone. The data are scarce but consistent with theoretical predictions of what a kilonova should look like. If this identification is correct, it is the first time we have directly witnessed the production of heavy, precious metals. We would like to observe more such events to better determine the amount of metals these explosions synthesize and whether they can account for all, or only part, of the abundance of gold, platinum and other heavy elements in the universe.

CHAOTIC COSMOS

OUR STUDY of the dynamic universe has just begun. Within a decade or so new automated telescopes such as the Zwicky Transient Facility coming online near San Diego, the Large Synoptic Survey Telescope being built in Chile and the Wide-Field Infrared Survey Telescope that NASA plans to launch into space will be able to scan most of the sky every few nights, discovering hundreds of times as many supernovae as we currently do. Meanwhile modern supercomputers are becoming capable of carrying out detailed three-dimensional simulations of these events, allowing us to visualize what may go on deep in the cores of exploding stars.

Data gathered in the coming years will challenge our theories about the many types of stellar deaths. Each scenario described here is physically plausible but unproved. With more observations of unusual supernovae, we hope to pin down which of these explosive possibilities are, in fact, realized in nature. Most likely, the universe will turn out to be stranger than we have imagined, revealing even more exotic phenomena than we have dreamed up so far.

Ultimately we will also be able to tell a richer narrative of the stuff that makes up our bodies and the world around us. The gold ring on your finger, for example, has a history that goes back beyond the time of your ancestors. That material probably first floated in the iron furnace of a massive star that faltered, collapsed and compressed into a dense neutron star. Much later, after maybe a billion years, the neutron star might have crashed into another compact star, spewing a cloud of radioactive waste out into space. That cloud, rushing at 60 million miles per hour, would have traveled more than 1,000 light-years across the galaxy, mixing with other gases along the way, until it eventually settled into the crust of planet Earth. Some time later people dug up that stellar rubble, shaped a ring and began to tell their own stories. SA

Daniel Kasen is an astrophysicist at the University of California, Berkeley, and at Lawrence Berkeley National Laboratory. His research focuses on developing new theoretical and computer models to explain the many types of stellar explosions in the universe.

MORE TO EXPLORE

Supernova Light Curves Powered by Young Magnetars. Daniel Kasen and Lars Bildsten in *Astrophysical Journal*, Vol. 717, No. 1, pages 245–249; July 1, 2010.

The Unprecedented 2012 Outburst of SN 2009ip: A Luminous Blue Variable Star Becomes a True Supernova. Jon C. Mauerhan et al. in *Monthly Notices of the Royal Astronomical Society*, Vol. 430, No. 3, pages 1801–1810; April 11, 2013.

Effect of a High Opacity on the Light Curves of Radioactively Powered Transients from Compact Object Mergers. Jennifer Barnes and Daniel Kasen in *Astrophysical Journal*, Vol. 775, No. 1, Article No. 18; September 20, 2013.

scientificamerican.com/magazine/sa

© 2017 Scientific American

EXTREME MOLECULES IN SPACE

Astrochemists are discovering many compounds in the cosmos that cannot exist on Earth

By Clara Moskowitz

IN BRIEF

The compounds that are found naturally on Earth—and that make up the huge diversity of materials on this planet—are just a fraction of those in the cosmos. **Astronomers have found** more than 200 types of molecules floating in space. **Researchers** have named some of the compounds they have detected in the universe and even re-created some here on Earth.

© 2017 Scientific American

HORSEHEAD NEBULA, as viewed by the Hubble Space Telescope.

© 2017 Scientific American

SOMETHING STRANGE WAS HIDING IN THE HORSEHEAD. THE NEBULA, NAMED for its stallionlike silhouette, is a towering cloud of dust and gas 1,500 light-years from Earth where new stars are continually born. It is one of the most recognizable celestial objects, and scientists have studied it intensely. But in 2011 astronomers at the Institute of Millimeter Radioastronomy (IRAM) and elsewhere probed it again. With IRAM's 30-meter telescope in the Spanish Sierra Nevada, they homed in on two portions of the horse's mane in radio light. They were not interested in taking more pictures of the Horsehead; instead they were after spectra—readings of the light broken down into their constituent wavelengths, which reveal the chemical makeup of the nebula. Displayed on screen, the data looked like blips on a heart monitor; each wiggle indicated that some molecule in the nebula had emitted light of a particular wavelength.

Every molecule in the universe makes its own characteristic wiggles based on the orientation of the protons, neutrons and electrons within it. Most of the wiggles in the Horsehead data were easily attributable to common chemicals such as carbon monoxide, formaldehyde and neutral carbon. But one spot in the Horsehead also had several small unidentified lines equally separated from one another in frequency. These represented a mystery—a molecule completely unknown to science.

Immediately after seeing the data, Evelyne Roueff of Paris Observatory and other chemists on the team started hypothesizing about what kind of molecule might create the signal. They concluded that the unknown species had to be a linear molecule—a compound whose atoms are arranged in a straight chain. Only a certain type of linear molecule would produce the spectral pattern the chemists were seeing. After working through lists of likely molecules, they hit on C_3H^+, propynylidynium. This molecular ion had never been seen before. In fact, there was no proof it existed at all. If it could form, it would be highly unstable. On Earth it would almost immediately react with something else to transform into a more settled species. But in space, where the pressure is low and molecules rarely run into anything else to bond with, C_3H^+ might just be able to survive.

Roueff and her colleagues studied whether the Horsehead Nebula contained the right ingredients and conditions to form the molecule. In 2012 they published a paper in *Astronomy & Astrophysics* concluding that the wiggle they saw likely belonged to C_3H^+. "I was relatively confident myself," Roueff says. "But it required about two to three years to convince everyone that we had the right identification."

At first, some skeptics contested the claim—if C_3H^+ had never been seen before, how could they be sure this was it? The clincher came in 2014, when researchers at the University of Cologne in Germany managed to create C_3H^+ very briefly in a laboratory. Not only did the feat prove that the molecule could exist, it also allowed scientists to measure the spectrum it produces when excited—the very same spectrum that showed up in the Horsehead. "It's very rewarding to find a new molecule that we did not really think about before," Roueff says. "When you are able to identify it through a chain of logic, it's like being a detective."

One alien molecule down, many, many more to go. The Horsehead Nebula is no aberration. Almost everywhere in the universe astronomers look—if they peer closely enough—they see unidentified spectral lines. The compounds we humans are familiar with, the species responsible for the huge diversity of materials on this planet, are just a fraction of those nature has created. And finally, after decades of work developing theoretical models and computer simulation techniques, along with laboratory experiments to reproduce new molecules, astrochemists are putting names to many of those unidentified lines.

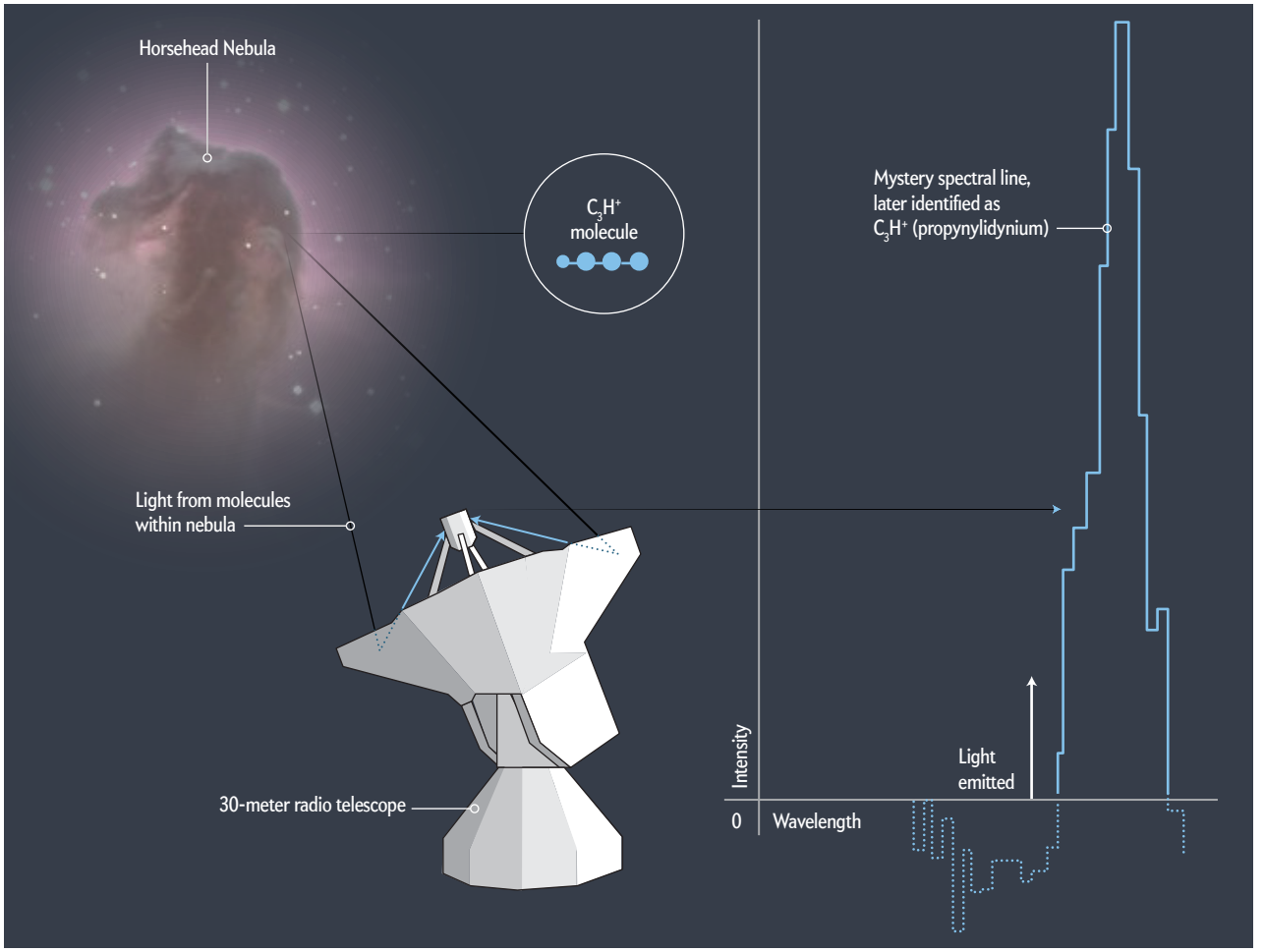
EMPTY SPACE

AS RECENTLY AS THE 1960S, most scientists doubted molecules could exist in interstellar space at all—the radiation there was thought to be too harsh for anything much beyond atoms and a few basic free radicals to survive. In 1968 physicist Charles Townes of the University of California, Berkeley, decided to look for molecules in space anyway. "I got the feeling that most of the Berkeley astronomers thought my idea was a little wild," Townes, a Nobel laureate who died in 2015, recalled in a 2006 account for the Astronomical Society of the Pacific. But Townes forged ahead and built a new amplifier for the six-meter antenna at Hat Creek Radio Observatory in California, which revealed the presence of ammonia in the Sagittarius B2 cloud. "How easy, and how exciting!" he wrote. "The news media as well as scientists began buzzing us."

In the years since, astronomers have found more than 200 types of molecules floating in space. Many are quite different from the species seen on the ground. "We usually do chemistry based on the conditions we have on Earth," says Ryan Fortenberry, an astrochemist at Georgia Southern University. "When we get

PRECEDING PAGES: NASA, ESA AND HUBBLE HERITAGE TEAM (AURA/STScI)

© 2017 Scientific American



ASTRONOMERS OBSERVING the Horsehead Nebula with a radio telescope in Spain saw the chemical signature of a mystery molecule. The telescope returned spectroscopic data—a line graph representing the intensity of light coming from the nebula across a range of wavelengths. Sharp rises in the light, such as the feature shown here, indicate the presence of a specific molecule whose chemical properties allow it to emit that particular wavelength of light. After much sleuthing, researchers were able to determine that this unidentified line feature comes from the never before seen compound C_3H^+, which exists only in space.

away from that paradigm, the chemicals that can be created are unbounded. If you can dream of a molecule, no matter how bizarre, there is a finite probability that over the eons of time and the immensity of space it has existed somewhere."

Space is literally an alien environment. Temperatures can be much, much hotter than on Earth (such as in the atmosphere of a star) and much, much colder (in relatively empty interstellar space). Likewise, the pressures (extremely high or low) are far from terrestrial. Consequently, molecules can form in space that would never emerge on our planet—and then they can stick around, even if they are highly reactive. "A molecule can go years and years before it bounces into another molecule in interstellar space," says Timothy Lee, an astrophysicist at NASA's Ames Research Center. "It might be in a region where there's no radiation, so even if it's not that stable, it can exist for a long time."

These space molecules, once identified, could teach us a lot. Some of them might prove beneficial if scientists can create them in labs and learn to exploit their properties. Other molecules may help explain the origins of the organic compounds that gave rise to life on Earth. And all of them stand to expand the bounds of what is possible for chemistry in the universe.

GAME-CHANGING TELESCOPES

IN THE PAST DECADE, as powerful new telescopes capable of observing faint spectral lines have come online, the search for alien molecules has accelerated. "It's a heyday for astrochemistry right now," says Susanna Widicus Weaver, who leads an astrochemistry group at Emory University. The data available even in 2015, she says, are a huge improvement from just a decade earlier when she finished her doctorate. NASA's high-altitude Stratospheric Observatory for Infrared Astronomy (SOFIA), mounted on the side of a Boeing 747SP, began observing in infrared and microwave light in 2010 and the European Space Agency's Herschel Space Observatory launched into orbit in 2009 to target the same wavelengths.

The real game changer, however, is the multinational Atacama Large Millimeter/submillimeter Array (ALMA), a constellation of 66 radio dishes inaugurated in 2013. At an altitude of about 5,200 meters on the Chajnantor Plateau, a Mars-like red expanse in

Graphic by Amanda Montañez

© 2017 Scientific American

1

SAGITTARIUS B2 CLOUD, where astronomers discovered ammonia in 1968.

2

ATACAMA Large Millimeter/submillimeter Array (ALMA) in Chile's Atacama Desert.

Chile's Atacama Desert, the driest place in the world, ALMA's matching antennas swivel in unison as observers collect light from cosmic objects. Extremely dark and clear skies nearly devoid of image-blurring moisture give the telescope unprecedented sensitivity and precision in wavelengths from infrared to radio. ALMA creates both a visual picture and a spectrum for every pixel of its images, producing tens of thousands of spectral lines in every field of view it observes. "It's amazing and it's overwhelming at the same time," Widicus Weaver says. "These data sets are so big that they often have to mail them to scientists on flash drives because they can't download them." The flood of data is providing a wealth of new spectral lines for astrochemists to mine. But like unidentified fingerprints at a crime scene, these lines are useless to scientists unless they can determine which molecules created them.

FINDING A LINK

TO MATCH MOLECULES to these lines, scientists can go in a few directions. As in the case of C_3H^+, astrochemists might start with clues from the spectrum to guess what molecule might be behind it. A technique called *ab initio* quantum chemistry (*ab initio* is Latin for "from the beginning") allows scientists to start from pure quantum mechanics—the theory that describes the behavior of subatomic particles—to calculate a molecule's properties based on the motions of the protons, neutrons and electrons in the atoms that compose it. On a supercomputer, scientists can run repeated simulations of a molecule, each time slightly adjusting its structure and the arrangement of its particles, and watch the results to find the optimal geometry of a compound. "With quantum chemistry we're not limited by what we can synthesize," Fortenberry says. "We're limited by the size of the molecules. We need large amounts of computational power to do the calculations."

Researchers can also look for hard evidence of new molecules by creating them in a lab and directly measuring their spectral features. A common technique is to start with a chamber of gas and run a current of electricity through it. An electron in the current might collide with a molecule of gas and break its chemical bond, giving rise to something new. Researchers keep the gas at very low pressure so that any chemicals that arise have a chance to hang out for a few moments before running into another molecule and reacting. Scientists will then shine various wavelengths of light through the chamber to measure the spectrum of whatever is inside. "You can get to the point where you've produced in the lab the same molecule that's occurring in space, but you don't necessarily know what the molecule is," says Michael McCarthy, a physicist at the Harvard–Smithsonian Center for Astrophysics. "So then you have to try to infer the elemental composition from a combination of different lab experiments with different samples."

In 2006 McCarthy and his colleagues created the negatively charged molecule C_6H^- and measured its spectrum. Soon afterward they found the same spectral features in the interstellar Taurus Molecular Cloud around 430 light-years away. Previous searches for negative molecules in space had come up empty, so many scientists doubted whether they existed in significant numbers. "It led us to a whole set of discoveries in which we were able to detect related molecules in the lab and then in space," McCarthy says. His team and others have now found C_6H^- in more than a dozen cosmic sources.

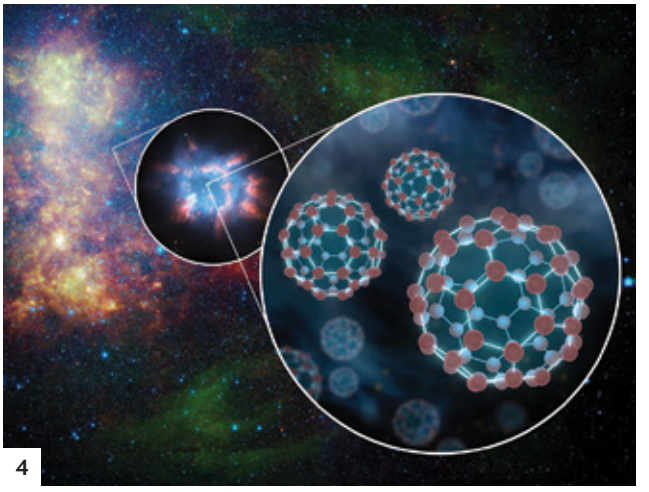
And in the 1980s scientists trying to make new chemicals produced the molecule argonium ($^{36}ArH^+$), a strange compound not normally found on Earth that combines hydrogen with the generally inert gas argon. In 2013 astronomers found argonium in space, first in the Crab Nebula and later in a distant galaxy via ALMA observations. Compounds based on noble gases form only under very specific circumstances; scientists think that in space, high-energy charged particles called cosmic rays slam into argon and knock electrons loose, enabling it to bond with hydrogen. For

ESO/APEX AND MSX/IPAC/NASA (1); MARTIN BERNETTI *Getty Images* (2)

© 2017 Scientific American

3

TAURUS MOLECULAR CLOUD, as seen by the Atacama Pathfinder Experiment (APEX) telescope in Chile.



4

BUCKYBALLS, conglomerations of 60 carbon atoms, were first created in laboratories on Earth and later discovered in space.

this reason, if scientists see argonium in a region of space, they can surmise that the place is flooded with cosmic rays. "It's a very specific indicator of these circumstances, which actually are very important in space," says Holger Müller of the University of Cologne, leader of the team behind the ALMA discovery.

A NEW WORLD OF MOLECULES

MANY OF THE MOLECULES LURKING in stars and nebulae are foreign in the extreme. To ask what they would look or feel like if you could hold them in your hand is nonsensical because you could never hold them—wthey would react immediately. If you *did* manage to make contact with them, they would almost certainly prove toxic and carcinogenic. Oddly, however, scientists have a rough idea of what some alien molecules would smell like: many detected so far belong to a class of compounds called aromatics, which are derived from benzene (C_6H_6) and were originally named for their strong odors.

Some new compounds reveal surprising atomic structures and share charge between atoms in unforeseen ways. Sometimes they challenge current theories of molecular bonding. A recent example is the molecule SiCSi, discovered in 2015 in a dying star, which is made of two silicon atoms and one carbon atom that are bonded in an unexpected way. The resulting molecule is somewhat floppy and produces a spectrum different from what simple theoretical models predict.

And space molecules may help answer one of the universe's most fundamental questions: How did life get started? Scientists do not know if amino acids, the building blocks of life, first arose on Earth or in space (and were then delivered to our planet by comets and meteorites). "The big question is, Do they form in molecular clouds as a star is forming," asks Widicus Weaver, "or do they form once you have a planet or some other chunk of rock where chemistry can occur on the surface?" The answer will determine whether it is likely that amino acids are plentiful in the universe and available to possibly seed life on the myriad exoplanets out there or whether the chemistry that sparked us was isolated to our own planetary cradle. Astrochemists have already spotted signs of amino acids in space as well as sequences of molecules that might have given rise to them.

Finally, some rare species could prove useful if they can be created in sufficient quantities and kept under controlled conditions. "The great hope of astrochemistry is that we can find molecules with completely new properties and apply those to problems here on Earth," Fortenberry says. An example is the soccer ball–shaped molecules "buckyballs." These large conglomerations of 60 carbon atoms first showed up in a lab in 1985 (and won their discoverers a Nobel Prize). Almost a decade later astronomers saw spectral features in interstellar gas that looked consistent with positively charged versions of buckyballs, and the connection was confirmed in 2015 when researchers matched those features to the spectrum of buckyballs created under spacelike conditions in the lab. "This molecule is now all over the galaxy and all over the universe," noted the late buckyball co-discoverer Harold Kroto, then a chemistry professor at Florida State University. Lately buckyballs have turned out to be not just a quirk found in space but a practical tool for nanotechnology, useful for strengthening materials, for improving solar cells and even for pharmaceuticals.

At this point astrochemists are still testing the shallow waters in the great sea of molecules out there in space. The finds they have already turned up are a reminder that our own small corner of the cosmos is just that—an insignificant, and not necessarily representative, sample of what is possible. Perhaps the species we are familiar with on Earth are in fact the exotic ones, and the buckyballs, the Horsehead Nebula, C_3H^+ and others still unknown are the ordinary stuff of the universe. SA

Clara Moskowitz is *Scientific American*'s senior editor covering space and physics. She has a bachelor's degree in astronomy and physics from Wesleyan University and a graduate degree in science journalism from the University of California, Santa Cruz.

MORE TO EXPLORE

Detection of a Noble Gas Molecular Ion, $^{36}ArH^+$, in the Crab Nebula. M. J. Barlow et al. in *Science* Vol. 342, pages 1343–1345; December 13, 2013.

The Interstellar Chemistry of C_3H and C_3H_2 Isomers. Jean-Christophe Loison et al. in *Monthly Notices of the Royal Astronomical Society,* Vol. 470, No. 4, pages 4075–4088; October 2017. Preprint available at arXiv: https://arxiv.org/abs/1707.07926.

scientificamerican.com/magazine/sa

ESO (3); NASA/JPL-CALTECH (4)

© 2017 Scientific American

© 2017 Scientific American

New techniques reveal how supermassive black holes shred entire stars

By S. Bradley Cenko and Neil Gehrels

How to Swallow a Sun

Illustration by Ron Miller

© 2017 Scientific American

AT THE HEART OF OUR MILKY WAY AND OF VIRTUALLY EVERY OTHER large galaxy lurks a deep cosmic mystery—a supermassive black hole. Squeezing millions to billions of times more mass than our sun into regions smaller in size than our solar system, these objects are so bizarre as to seem almost mystical. No one yet understands exactly how nature has managed to compress so much matter into such small spaces. What is clear is that the hidden gravitational hands of supermassive black holes reach out to shape their surrounding galaxies in profound yet subtle ways. By studying the growth and behavior of these ghostly black holes, scientists hope to unlock the secrets of how galaxies themselves are born and evolve.

The trouble is that because they emit no light, supermassive black holes spend most of their time dormant and invisible to us. They come to life only when they feed, but their meals are surprisingly rare—most of the gas, dust and stars swirling around them reside in stable orbits and will never be consumed. Yet they are always hungry, and whenever something sizable does happen to fall in, the resulting feeding frenzies can be seen from far, far away.

For most of the past half a century, scientists have primarily observed just one variety of feasting black holes: quasars. Discovered by astronomer Maarten Schmidt in 1963, quasars are the ultraluminous centers of active galaxies that can be seen all the way to the outskirts of the visible universe, each shining brighter than billions of suns. They are thought to occur when massive clouds of gas and dust plunge onto a supermassive black hole for hundreds of thousands or millions of years, compressing, heating and glowing as they circle the black hole's maw. Quasars, however, are less than ideal objects of study. They are extreme events that are typically quite distant and relatively rare, and they constitute a small portion of any supermassive black hole's lifetime. They thus provide a limited view, leaving astronomers blind to how supermassive black holes more routinely feed and grow in the local universe. Researchers have also studied supermassive black holes by clocking the speeds of stars whizzing around them, but these measures work solely for objects very close to home—in the Milky Way or one of our immediate galactic neighbors—where current telescopes can resolve individual stars.

In 1988 astronomer Martin Rees proposed a third way to study supermassive black holes—a method that has recently begun to bear fruit. Instead of watching for the steady glow of quasars or the speeds of orbiting stars, astronomers can look for brief, bright flares of light from a black hole's vicinity. Called tidal disruption events (TDEs), these outbursts occur when a supermassive black hole consumes an unlucky star. Unfolding over months rather than millennia, they allow researchers to track the feeding process from beginning to end and are bright enough to be observed in galaxies both near and far.

IN BRIEF

A star passing too close to a supermassive black hole will be ripped apart by tidal forces and consumed.

New wide-field telescopes are allowing astronomers to study these cataclysmic events in fine detail.

The result is a dramatic improvement in our understanding of how black holes feed, how they can accelerate material to velocities approaching the speed of light and how their growth shapes their surrounding galaxies.

© 2017 Scientific American

COSMIC BEACONS CALLED QUASARS are produced by supermassive black holes feasting on gas (*above*) but are too rare, distant and sluggish to wholly reveal how such giant black holes feed. More nuanced detail can come from watching black holes snack on entire stars.

ILLUSTRATION COURTESY OF NASA AND JPL/CALTECH

HOW TO DESTROY A STAR

A TIDAL DISRUPTION EVENT is far more dramatic than the gentle tides that can wash your beach towels away at the seashore, but it is not all that different in principle. Earth's tides are chiefly caused by the gravitational tug of the moon, which pulls more strongly on the side of Earth closest to it. This difference between the moon's gravitational pull on the far and near sides of Earth is called a tidal force. That tidal force raises a bulge, or high tide, both on the moon-facing side of Earth and, somewhat paradoxically, on the opposite face as well, which also creates a

© 2017 Scientific American

corresponding low tide, oriented 90 degrees from the Earth-moon axis. When a star is in the vicinity of a supermassive black hole—perhaps pushed there by the gravitational nudging of another nearby star—the intense tidal forces can rip it to shreds.

The details of the star's demise depend on the sizes of both the infalling star and the supermassive black hole. A small, dense object, such as a white dwarf star, will be far more resilient against tidal forces than a larger, puffier, sunlike star—similar to how a bowling ball would be harder to rip apart than a mass of cotton candy. The very largest supermassive black holes, those containing billions of solar masses, are too big to easily create TDEs—they swallow stars whole before the tidal force becomes large enough to tear the stars apart. The tidal forces around a black hole with millions of solar masses, in contrast, will tear apart most stars that approach within about 50 million kilometers—roughly the distance of Mercury from the sun.

As spectacular as the wholesale disassembly of a star may seem, this is just the beginning of the fireworks display. After the initial disruption, the stellar debris will spread out and gradually diverge from the star's original orbital path. Basic orbital mechanics dictates that roughly half of the debris will be expelled as long filaments of material streaming from the vicinity of the black hole, whereas the other half will loop back around to form an accretion disk—a structure of spiraling rings slowly funneling into the black hole. When the disk's material falls in, it accelerates to nearly the speed of light and glows as gravitational and frictional forces compress and heat it to temperatures approaching 250,000 degrees Celsius. Over a span of weeks or months a typical TDE will cause a previously dormant and invisible black hole to briefly outshine all the stars in its galaxy.

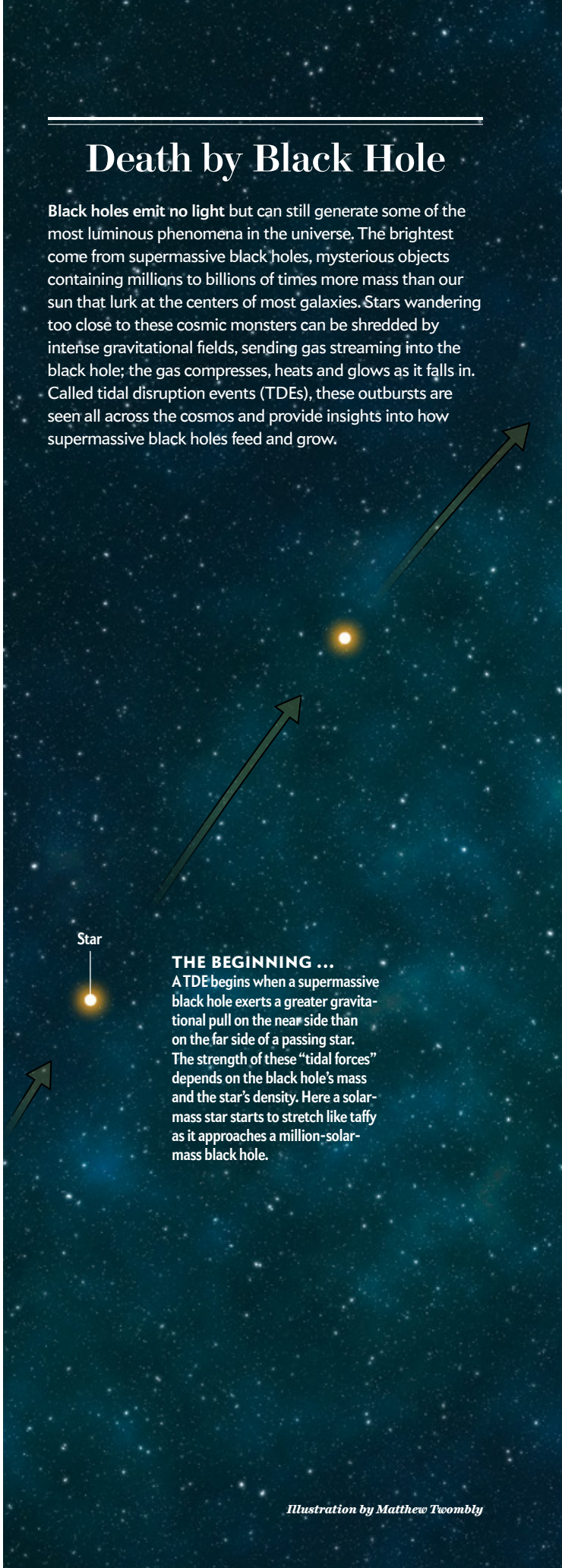
THE FIRST DISCOVERIES

ALTHOUGH THEORISTS first predicted TDEs decades ago, astronomers did not detect any until the 1990s and early 2000s. This delay was in part caused by the rarity of TDEs—they occur once every 100,000 years in a galaxy like the Milky Way, according to estimates. They can also be hard to see. Simple theoretical models suggested that the glow of a TDE's accretion disk should peak in the so-called soft x-ray or far ultraviolet parts of the electromagnetic spectrum—wavelengths that are difficult to access from the ground because of interference from interstellar dust and Earth's atmosphere. Those same models also suggested that astronomers could use a TDE to make relatively precise estimates for the mass of its accompanying black hole—a critical data point for learning exactly how a black hole's size alters its behavior and effects on its galactic surroundings. To measure a black hole's mass, astronomers can simply time how long an accompanying TDE takes to reach peak brightness (which reveals how fast an accretion disk forms and feeds the black hole). Because TDEs are so bright, they allow researchers to pinpoint masses for a wider range of supermassive black holes than any other known phenomenon.

The first TDE candidates were found in data from the ROSAT x-ray and Galaxy Evolution Explorer ultraviolet space telescopes. They appeared as weeks- to months-long flaring

Death by Black Hole

Black holes emit no light but can still generate some of the most luminous phenomena in the universe. The brightest come from supermassive black holes, mysterious objects containing millions to billions of times more mass than our sun that lurk at the centers of most galaxies. Stars wandering too close to these cosmic monsters can be shredded by intense gravitational fields, sending gas streaming into the black hole; the gas compresses, heats and glows as it falls in. Called tidal disruption events (TDEs), these outbursts are seen all across the cosmos and provide insights into how supermassive black holes feed and grow.



THE BEGINNING ...
A TDE begins when a supermassive black hole exerts a greater gravitational pull on the near side than on the far side of a passing star. The strength of these "tidal forces" depends on the black hole's mass and the star's density. Here a solar-mass star starts to stretch like taffy as it approaches a million-solar-mass black hole.

Illustration by Matthew Twombly

© 2017 Scientific American

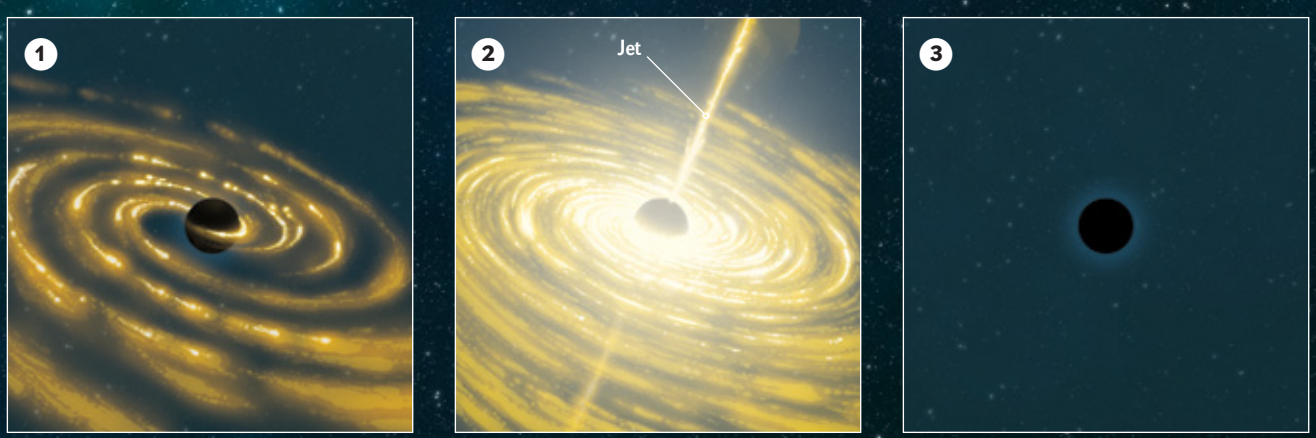
Black hole

... OF THE END

The tidal forces increase as the star gets closer to the black hole, until they overcome the star's self-gravity and rip it apart into arcing filaments of gas. Half the filaments are flung away from the black hole, never to return. The rest swirls into decaying orbits, forming a white-hot accretion disk that funnels the stellar debris into the black hole. Astronomers detect TDEs via the glowing accretion disks left in their wake.

A VIEW LIKE NO OTHER

A TDE is the only known way for astrophysicists to witness a supermassive black hole awaken, feast on a meal and return to quiescence. By tracking how long an accretion disk takes to form (1), reach peak brightness (2) and fade (3), scientists can gauge the size of the consumed star, as well as the black hole's mass and spin. As the black hole feeds, researchers can also monitor shock waves in its accretion disk and the creation of relativistic jets—streams of particles launched at near light speed from a black hole's poles. No other cosmic event offers such detailed views of these extreme, dynamic processes around black holes.



© 2017 Scientific American

events sourced to the centers of previously dormant galaxies. As the first manifestations of this long-predicted phenomenon, these discoveries were especially important for establishing an entirely new field of study. Because they were predominantly found in old data, however, astronomers were unable to study them across multiple wavelengths in real time to unlock their deepest secrets. To catch TDEs as they happened, astronomers would either have to be very lucky or able to continually survey mammoth swaths of sky.

As luck would have it, in the past decade steady progress in data storage and sensors has made such ambitious surveys possible. A high-end optical camera can now image one square degree or more of sky in a single snapshot, a situation akin to suddenly seeing the heavens with a panoramic lens after years of studying them by looking through a straw. By repeatedly surveying large areas of the sky and digitally combining the resulting images to tease out faint, temporary features, astronomers are now able to more easily discover and study TDEs and a host of other transient astrophysical phenomena. These new wide-field surveys, with names such as the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS), the Palomar Transient Factory (PTF) and the All-Sky Automated Survey for Supernovae (ASAS-SN), were designed largely to identify supernovae and asteroids, but they can do much more. Because they can image millions of galaxies each night, they are also sensitive to more exotic transients such as TDEs.

BY STUDYING TIDAL DISRUPTION EVENTS, ASTRONOMERS ARE LEARNING NOT ONLY ABOUT BLACK HOLES BUT ALSO ABOUT STARS BEING RIPPED APART BILLIONS OF LIGHT-YEARS AWAY.

NEW QUESTIONS FOR A NEW ERA

IN 2010, SHORTLY AFTER Pan-STARRS's on-sky debut, a team led by astronomer Suvi Gezari discovered a TDE called PS1-10jh, which took place around a black hole of perhaps two million solar masses in a galaxy some 2.7 billion light-years from Earth. Because this TDE was noticed soon after the data were collected, Gezari and her colleagues were, for the first time, able to watch it unfold in subsequent optical and ultraviolet observations. What they found was shocking.

Based on careful measurements of its spectrum, this particular TDE appeared to be much too cold. At about 30,000 degrees C, it was more than eight times cooler than predicted by most basic theories of accretion disks. Moreover, rather than fading over weeks as its accretion disk cooled and dissipated, PS1-10jh maintained a constant temperature for many months after the initial discovery. Strangest of all, Pan-STARRS had detected signs of ionized helium in the afterglow—something that could be produced only by temperatures in excess of 100,000 degrees C. And although the TDE appeared to be helium-rich, it also seemed to be devoid of hydrogen, the most abundant element in the universe and the main ingredient of stars. Theorists got to work examining what could produce such confounding results.

To explain PS1-10jh's lack of hydrogen, the Pan-STARRS team suggested that the disrupted star had lost its thick envelope of hydrogen at some point in the past, possibly in a previous interaction with the black hole, leaving only the star's helium-rich core to feed the observed accretion disk. But this alone could not account for the TDE's curious thermal discrepancies—its surprisingly low temperature and, in contradiction, its abundance of helium ionized by much higher temperatures. To solve that mystery, other theorists have postulated that PS1-10jh's black hole–hugging accretion disk was not actually being directly observed. Instead astronomers must have seen a surrounding veil of gas much farther out from the black hole that absorbed the intense radiation produced by the accretion disk and reemitted it at lower temperatures. This veil would have the added bonus of clarifying the apparent absence of hydrogen without the need for an exotic, helium-rich core as the TDE's progenitor. Given the right temperature and a rather high density, such a veil could potentially mask the presence of hydrogen, hiding it in plain sight.

The only trouble was that a thick veil of gas would not be stable at the required distance from the galaxy's central black hole—over time the gas would either fall into the black hole or dissipate to invisibility. The murky origins of this material are still the subject of intense debate and study but, broadly speaking, fall into two possibilities, both related to the dynamics of a feeding black hole. As the remains of a disrupted star loop around a black hole to form a growing accretion disk, shock waves can ripple outward well away from the disk to forestall some of the outlying debris from immediately falling in, creating a temporary screen of material. Alternatively, a newborn TDE accretion disk might initially funnel so much material inward that it briefly exceeds a black hole's capacity to feed, forming transient winds or outflows just outside the black hole that push stellar debris out past the accretion disk to much larger distances.

As astronomers sorted through these muddled possibilities for PS1-10jh and other newfound TDEs that soon followed, one thing became perfectly clear: TDEs were a much more complicated phenomenon than anyone had previously appreciated. Yet the biggest surprise was still to come.

A SHOCK FROM SWIFT

IN THE PREDAWN HOURS of March 28, 2011, an automated alert buzzed on the pagers and cell phones of a dedicated team of astronomers around the world. The Swift satellite had just detected a pulse of high-energy radiation from the depths of space. Built by NASA in partnership with research institutions in Italy and the U.K., Swift is an agile space telescope, designed to study all types of exploding objects across the sky. But its main targets are gamma-ray bursts, or GRBs—cataclysmic stellar explosions that are the most luminous astrophysical events in the universe. Whenever a torrent of gamma rays trickles into Swift's sensors, the telescope quickly reorients to observe the source in x-rays and optical light and phones home to trigger a complex chain of events on Earth. Within moments of an alert from Swift, astronomers scramble to commandeer the world's largest and most powerful telescopes to look for any emberlike afterglow associated with a GRB before it fades from view forever. Since its launch in 2004, Swift has discovered more than

© 2017 Scientific American

1,000 GRBs, but this particular event, later dubbed Swift J1644+57, would prove to be unlike anything the satellite had seen before.

As their name suggests, GRBs tend to be brief, typically lasting anywhere between a fraction of a second and a few minutes. When we turned our telescopes to Swift J1644+57 on that early March morning, we expected to see the standard, fading afterglow of a short-lived GRB. Instead we observed bright, erratic gamma-ray flaring that lasted for a day, followed by months of intense but fading x-ray emissions. Soon we had traced the blast to a galaxy some 3.8 billion light-years away in the constellation of Draco. One of our colleagues, Joshua S. Bloom of the University of California, Berkeley, suggested we had witnessed a TDE and correctly predicted that this particular gamma-ray source would be found at the galaxy's center—the stomping ground of supermassive black holes. But whereas all previous TDEs had been detected at longer, lower-energy wavelengths where observers saw the thermal emission from the accretion disk of a shredded star, this was something wholly different.

How could a TDE produce gamma rays? The best answer we could come up with is that black holes are messy eaters. A black hole would devour most of a disrupted star's gas, locking it away forever behind an event horizon (the boundary past which the black hole's gravitational pull is so great that not even light can escape). But all black holes probably spin, and this spin can push a few percent of the disrupted star's total gas toward the black hole's poles, outside the event horizon, where the gas is accelerated and ejected as collimated beams of particles moving at nearly the speed of light. The fast-moving beams emit gamma rays and x-rays as they hurtle through the cosmos. Apparently Swift had happened to find itself in the path of Swift J1644+57's beam. It was a lucky catch—not all TDEs appear to power such relativistic outflows, and most of those that do likely miss our line of sight.

With the detection of Swift J1644+57 as inspiration, the Swift team began a concerted search for more events. As of early 2017, two more TDEs emitting gamma-ray jets have been found. These rarest and most intense death cries of stars are a novel way to study one of the foremost research topics in modern high-energy astrophysics—the creation and behavior of relativistic particle jets.

THE DEATH OF WORLDS

WHETHER THROUGH the thermal emissions from accretion disks of stellar debris or rather with the gamma rays pouring off a star-devouring black hole's relativistic jets, TDEs offer a new window on the behavior and evolution of supermassive black holes and their surroundings. Most important, unlike the far larger and longer-duration jets and accretion disks of quasars, which are produced by immense clouds of gas chaotically plummeting onto a supermassive black hole across very long timescales, TDEs are brief, clean events that are more easily studied. No human being will ever live long enough to witness the complete life cycle of a single quasar, but already astronomers have found and studied more than 20 TDEs from beginning to end. And in the details of those stellar catastrophes, they have glimpsed tantalizing quirks begging for further study. By precisely measuring the fluctuating flares from TDEs, astronomers are learning not only about black holes but also about the detailed compositions and internal structures of stars being ripped apart billions of light-years away.

They may even eventually learn about a star's companions, too—planets swallowed by black holes. Each flickering flash from a faraway galactic center could signal the deaths of entire worlds. Surveys of stars in our own galaxy have found that almost every one harbors planets; planets probably accompany most, if not all, of the stars in other galaxies as well, including those that suffer TDEs. Even if they were not directly consumed, planets could still find themselves in the path of the transient relativistic jets produced by some TDEs, which extend light-years beyond their black hole sources. Life in any planetary systems unlucky enough to be struck by such a beam would be rapidly extinguished. Someday astronomers may witness a TDE right in our cosmic backyard, when the four-million-solar-mass black hole quiescently lurking at the gas-impoverished center of our very own Milky Way flares to life as it consumes some errant star. It would be very bright but also very safe because we are too far from the galactic center for the TDE's most dangerous effects to reach us here.

The advent of even more powerful surveys heralds a new era of TDE discoveries. The Large Synoptic Survey Telescope (LSST), an eight-meter telescope currently being constructed in Chile with a field of view covering 10 square degrees of sky, will by itself uncover thousands of these outbursts within a decade of its debut. In some ways, the most challenging aspect of LSST science will be sifting through the overwhelming number of transient discoveries. Planned radio observatories such as the Square Kilometer Array being built in Australia and South Africa are particularly well suited to identify relativistic jets, even if such jets are "off axis," meaning they are not beamed directly along our line of sight. In the not too distant future, astronomers may have assembled a TDE catalog with thousands on thousands of entries, more than any individual could study in a lifetime, shedding new light on the masses and behaviors of those elusive hungry ghosts, the otherwise inaccessible supermassive black holes that reside in the hearts of galaxies all across the universe. From that rich, ever growing corpus of knowledge, we can only dream of what further revolutionary discoveries may come. SA

S. Bradley Cenko is a research astrophysicist at NASA's Goddard Space Flight Center, where he serves as a deputy project scientist for the Swift mission. He works with wide-area ground- and space-based surveys to study the time-variable universe.

Neil Gehrels was chief of the Astroparticle Physics Laboratory at NASA's Goddard Space Flight Center, as well as principal investigator of the Swift mission and project scientist for the Wide-Field Infrared Survey Telescope (WFIRST) and the Fermi Gamma-ray Space Telescope. He passed away in February 2017.

MORE TO EXPLORE

Tidal Disruption of Stars by Black Holes of 10^6–10^8 Solar Masses in Nearby Galaxies. Martin J. Rees in *Nature,* Vol. 333, pages 523–528; June 9, 1988.

A Possible Relativistic Jetted Outburst from a Massive Black Hole Fed by a Tidally Disrupted Star. Joshua S. Bloom et al. in *Science,* Vol. 333, pages 203–206; July 8, 2011.

An Ultraviolet–Optical Flare from the Tidal Disruption of a Helium-Rich Stellar Core. S. Gezari et al. in *Nature,* Vol. 485, pages 217–220; May 10, 2012.

PS1-10jh Continues to Follow the Fallback Accretion Rate of a Tidally Disrupted Star. S. Gezari et al. in *Astrophysical Journal Letters,* Vol. 815, No. 1, Article No. L5; December 10, 2015.

scientificamerican.com/magazine/sa

© 2017 Scientific American

FROM THE SCIENTIFIC AMERICAN BOOKS IMPRINT

To know whether life exists beyond Earth, we must come to terms with our own significance in the universe. Are we uniquely special or merely mediocre?

By Caleb Scharf

COSMIC (IN)SIGNIFICANCE

WE ALL RESIDE ON A SMALL PLANET ORBITING A SINGLE, MIDDLE-AGED star that is one of some 200 billion stars in the great swirl of matter that makes up the Milky Way galaxy. Our galaxy is but one of an estimated several hundred billion such structures in the observable universe—a volume that now stretches in all directions from us for more than 270,000,000,000,000,000,000,000 (2.7×10^{23}) miles.

By any paltry human standard, this is an awful lot of stuff and an awfully large amount of room. Our species has sprung into existence within the barest instant of this universe's enormously long span of history, and it looks like there will be an even longer future that may or may not contain us. The quest to try to find our place, to discover our relevance, can seem like a monumental joke. We must be appallingly silly to imagine we can find any importance for ourselves at all.

Yet we are trying to do just that, despite our apparent mediocrity, which became evident when Renaissance scholar Nicolaus Copernicus decentralized Earth from the solar system around 500 years ago. His idea has been one of the greatest scientific guides for the past few hundred years and a critical signpost

Adapted from The Copernicus Complex: Our Cosmic Significance in a Universe of Planets and Probabilities, *by Caleb Scharf, by arrangement with Scientific American/Farrar, Straus and Giroux, LLC (US) and Penguin Press (UK).* *Copyright © 2014 by Caleb Scharf.*

© 2017 Scientific American

Illustration by Tim Bower

© 2017 Scientific American

on our journey to discern the underlying structure of the cosmos and the nature of reality.

In our efforts to assess our significance, we face a conundrum: Some discoveries and theories suggest life could easily be ordinary and common, and others suggest the opposite. How do we begin to pull together our knowledge of the cosmos—from bacteria to the big bang—to explain whether or not we are special? And as we learn more about our place in the universe, what does it all imply for our efforts to find out if there are other living things out there? How do we take the next steps?

WHAT WE KNOW

IN THE 1600s TRADESMAN and scientist Antony van Leeuwenhoek used his hand-built microscopes to become the first human to see bacteria, a journey that took him into the alien world of the microcosm. In that remarkable descent, sliding down the ladder of physical dimensions, into the thriving universe within us, was one of the first clues that the components of our bodies, our arrays of molecular structures, exist at one extreme end of a spectrum of biological scales. Until van Leeuwenhoek's moment of surprise, I doubt that humans had the opportunity to think about this fact in anything more than a superficial way.

There are organisms on Earth that are physically larger and more massive than we are—just look at whales and trees. Yet we are much closer to the upper limit of scale than we are to the microscopic end of life's spectrum. The smallest reproducing bacteria are a couple of hundred-*billionths* of a meter across; the smallest viruses are 10 times as small as that. The human body is roughly 10 million to 100 million times larger than the simplest life we know of.

Among warm-blooded terrestrial mammals we are also on the large side but not quite at the extreme top. At the opposite end of scale, the smallest of our kin are the pygmy shrews, diminutive scraps of fur and flesh barely two grams in weight. They exist at the edge of feasibility, their bodies endlessly leaking heat that they can barely compensate for by voracious eating. But most mammals are closer to this size than to our size: so much so that the global average body weight of the mammalian population is 40 grams, or less than 1.5 ounces. Our complex-celled, intelligent bodies are at the boundary of the upper extremes, with comparatively few mammalian types bigger than us.

It is an undeniable observation that we exist at this border, this interface between the complex diversity of the biologically small and the limited options of the biologically large. Consider, too, our planetary system. It is unusual in certain respects. Our sun is not one of the most numerous types of star (most of which are less massive), our orbits are at present more circular and rather more widely spaced apart than most exoplanetary systems, and we do not count a super-Earth among our planetary neighbors. Such a world, a few times more massive than Earth, is represented in at least 60 percent of all systems but not our solar system. If you were an architect of planetary systems, you would consider ours to be an outlier, a little bit off from the norm.

Some of these characteristics stem from the fact that our solar system has escaped wholesale dynamical rearrangement, compared with the majority of planetary systems. This does not mean that we are assured a quiet and peaceful future—state-of-the-art gravitational simulations indicate that a few hundred million years along, a more chaotic period could overtake our system. And another five billion years into the future the sun will inflate with the onset of a spasmodic old age and quite drastically revise the properties of its array of planets. All indications are that today we also live at an interface or border in time, a transition between a period of stellar and planetary youth and one of encroaching decrepitude. Our existence in this period of relative calm is, in retrospect, not so surprising. As with so many other aspects of our circumstances, we live in a temperate place, not too hot or cold, not too chemically caustic or chemically inert, neither too unsettled nor too unchanging.

It is also now apparent that this astrophysically calm neighborhood extends well beyond our local galaxy. In terms of the universe as a whole, we exist in a period that is many times more ancient than the fast tumult of the young, hot cosmos. Everywhere the production of stars is slowing down. Other suns, and their planets, are forming at an average rate that is barely 3 percent of that 11 billion to eight billion years ago. The stars are slowly beginning to go out across the universe. And in grand cosmological terms, only six billion or five billion years ago the universe was decelerating from the big bang. Now we are again in a period of gentle transition. Dark energy, stemming from the vacuum itself, is accelerating the growth of space, helping to quash the development of larger cosmic structures. But this means that life is ultimately condemned to a distant future of bleak isolation within an increasingly indecipherable universe.

Put all these factors together, and it is clear that our view of our inner and outer cosmos is highly constrained. It is a view from a narrow perch. Indeed, our basic intuition for random events and our scientific development of statistical inference might have been different under other circumstances of order or disorder, space and time. And the very fact that we are far isolated from any other life in the cosmos—to the extent that we have not spotted or stumbled across it yet—profoundly impacts the conclusions we can draw.

IN BRIEF

Earth orbits one of the hundreds of billions of stars in our galaxy, which in turn is one of hundreds of billions in the observable universe. This apparent insignificance fits with the Copernican principle that our planet is not the center of the cosmos but simply a mediocre member of a mediocre solar system.
Meanwhile there are reasons to think Earth and its life are special, perhaps even singular. Some evidence comes from the details of our planetary circumstances, as well as from the observation that certain fundamental constants of nature appear to be "fine-tuned" for life's existence.
Scientists must reconcile these conflicting ideas to understand where we fit—and whether we are alone—in the cosmos.

© 2017 Scientific American

OUR SPECIES HAS SPRUNG INTO EXISTENCE WITHIN THE BAREST INSTANT OF THIS UNIVERSE'S ENORMOUSLY LONG HISTORY, AND IT LOOKS LIKE THERE WILL BE AN EVEN LONGER FUTURE THAT MAY OR MAY NOT CONTAIN US. THE QUEST TO FIND OUR PLACE, TO DISCOVER OUR RELEVANCE, CAN SEEM LIKE A MONUMENTAL JOKE.

DEDUCTIONS

MUCH OF THE EVIDENCE we have supports the basic Copernican view that we are mediocre. Yet at the same time, there are specifics about our environment that say otherwise. Some of these qualities have led to the so-called anthropic principle, the observation that certain fundamental constants of nature appear "fine-tuned" in a way that causes the underlying properties of the universe to be balanced near a boundary that enables Earth and its life to exist. A little too far to either side, and the nature of the cosmos would be radically different. Tweak the relative strength of gravity, and either no stars form, no heavy elements are forged—or huge stars form and are quickly gone, leaving nothing of any import in their wake, no descendants, no pathway to life. Similarly, alter the electromagnetic force, and the chemical bonds between atoms would be too weak or too strong to build the diversity of molecular structures that allows such incredible complexity in the cosmos.

What do we make of all the contradictions? I would argue that the facts are pushing us toward a new scientific idea about our place in the cosmos, a departure from both the Copernican and anthropic principles, and I think it is well along the road to becoming a principle in its own right. Perhaps we could call it a cosmo-chaotic principle, the place between order (from the original Greek *kosmos*) and chaos. Its essence is that life, and specifically life like that on Earth, will always inhabit the border or interface between zones defined by such characteristics as energy, location, scale, time, order and disorder. Factors such as the stability or chaos of planetary orbits, or the variations of climate and geophysics on a planet, are direct manifestations of these characteristics. Too far away from such borders, in either direction, and the balance for life tips toward a hostile state. Life like us requires the right mix of ingredients, of calm and chaos—the right yin and yang.

Proximity to these edges keeps change and variation within reach but not so close that they overwhelm a system constantly. There are obvious parallels to the concept of a Goldilocks zone, which proposes that a temperate cosmic environment for a planet around a star exists within a narrow range of parameters. But for the existence of life, the hospitable zone may be much more dynamic—it need not be fixed in space or time. Rather it is a constantly drifting, twisting, flexing, multiparameter quantity, like the paths traced by a dancer's limbs.

If it is a universal rule that life exists only under these circumstances, it raises some intriguing possibilities about our cosmic significance. Unlike strict Copernican ideas, which stress our mediocrity and therefore suggest an abundance of similar circumstances across the cosmos, the notion that life requires a varying and dynamic alignment of parameters narrows the options. The opportunities for life implied by this new view also differ from anthropic ideas, which at their most extreme predict as little as one sole occurrence of life across all space and time. Instead this new rule actually identifies the places where life should occur and the potential frequency with which it does. It specifies the fundamental characteristics necessary for life within a virtual space of many waltzing parameters—it maps out the fertile zones.

Such a rule about life does not necessarily make living things some special part of reality. Biology may be the most complicated physical phenomenon in this universe—or in any amenable universe. But that is possibly as special as it gets: a particularly intricate natural structure that arises under the right circumstances, between order and chaos. And this conceptualization of where life fits into the grand scheme of nature leads directly to a way to resolve the conundrum between the persuasive, but unresolved, arguments that life must be abundant and that it is exquisitely rare. SA

Caleb Scharf is director of the multidisciplinary Astrobiology Center at Columbia University and author of *Gravity's Engines: How Bubble-Blowing Black Holes Rule Galaxies, Stars, and Life in the Cosmos* (Scientific American/Farrar, Straus and Giroux, 2012).

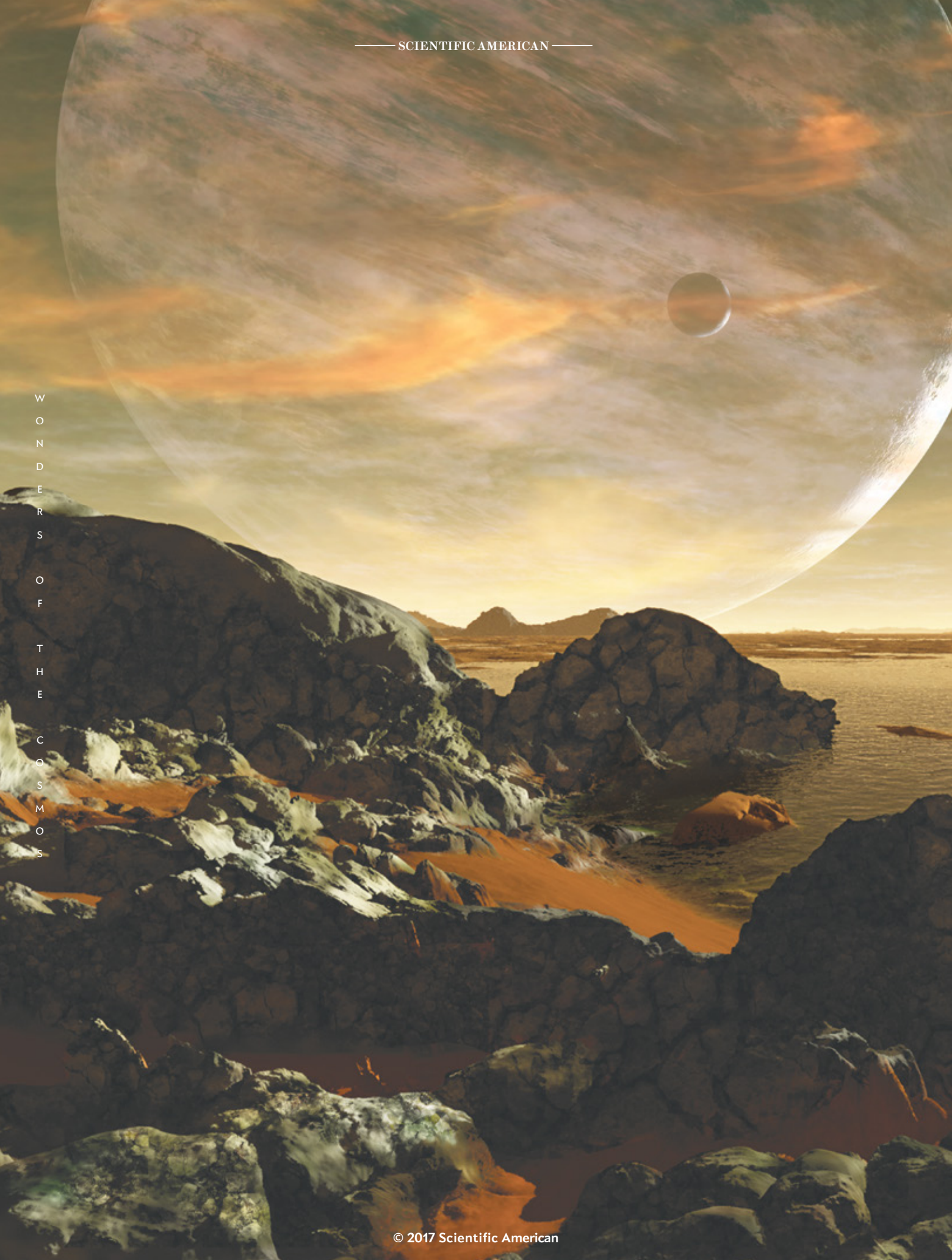
MORE TO EXPLORE

Just Six Numbers: The Deep Forces That Shape the Universe. Martin Rees. Basic Books, 2001.

Caleb Scharf's **Life, Unbounded** blog: http://blogs.ScientificAmerican.com/life-unbounded

scientificamerican.com/magazine/sa

© 2017 Scientific American



© 2017 Scientific American

© 2017 Scientific American

The Search for Life on Faraway Moons

Moons orbiting distant exoplanets may account for most of the habitable locales in the galaxy. If only we could find them

By Lee Billings

Illustrations by Ron Miller

WE NOW KNOW OF THOUSANDS OF PLANETS ORBITING OTHER STARS. In all likelihood, hundreds of billions more call the Milky Way home. Many of the known "exoplanets" are large, gaseous worlds like Jupiter or Neptune—hostile places for life. But like those giants of our solar system, distant exoplanets may also have large moons. And if they do, moons—not planets—may be the most common home for life in the universe.

In recent years one frontier of the search for moons of exoplanets—exomoons—has been in the basement of the Harvard-Smithsonian Center for Astrophysics, inside a gloomy room lined with computers in wire-mesh cages. Raising his voice over the mechanical whine of the cooling fans, British astronomer David Kipping remarks that nearly all of this computing power is currently devoted to analyzing a single planet, Kepler-22b, which orbits a sunlike star some 600 light-years away from Earth. The distant world is named for NASA's planet-hunting Kepler space telescope, which first spotted it. Kipping's hope is that on closer inspection, the data that first revealed Kepler-22b's presence may also divulge the subtler signals of lunar companions. He calls his project the Hunt for Exomoons with Kepler, or HEK.

Kipping's project is arguably the most advanced exomoon hunt today. The intense computing power is necessary, Kipping says, because even the largest conceivable exomoon would leave a vanishingly faint signal in the data. Because of this, he intensively searches for evidence of exomoons around only a subset of carefully selected targets. He may not find as many exomoons as he would with a quick search of lots of targets, but "I'm not sure I'd believe those results," he says. "Our goal is nice, clean, solid detections that everyone can agree on."

He has reason to be circumspect. Any claim of an exomoon discovery would be controversial, not only because the work is difficult but also because the find potentially has profound implications. For instance, Kipping explains, Kepler-22b resides in its star's habitable zone, the region where liquid water could exist. The planet is so large it is likely to be an inhospitable, gas-shrouded orb rather than a rocky, terrestrial world like Earth. If, however, Kepler-22b has a massive lunar companion, that moon might be a pleasant place to live and a possible target for future astronomical searches for extraterrestrial life and intelligence.

"Moons could be habitable," he says. "And if that's true, there's a hell of a lot more opportunities for life out there than anyone has previously appreciated."

MAKING MOONS

MANY ASTRONOMERS (as well as science-fiction authors) had long assumed that other planetary systems would mirror our own, with bountiful icy moons orbiting cold, giant worlds, similar to the arrangements we see around Jupiter and Saturn. With the first exoplanet discoveries of the 1990s, however, new possibilities arose; planet hunters began finding gas-giant planets that, after forming in the outer dark, somehow migrated in to closer, hotter orbits around stars. Some even occupied their stars' habitable zones. Such positioning raised the question: Might some moons around those warm giants have rocky compositions, protective atmospheres and oceans like on Earth?

Three researchers at Pennsylvania State University—Darren Williams, Jim Kasting and Richard Wade—were the first to study in detail how feasible it would be for an exomoon to possess an Earth-like environment. Their study, published in 1997 in *Nature,* asked how large a habitable-zone moon must be to maintain a substantial atmosphere and liquid water on its surface. "We found that moons smaller than Mars, about a tenth the mass of Earth, couldn't hold on to an atmosphere for more than a few million years," Williams says. Below that threshold, a moon would not generate enough gravitational force to retain a substantial atmosphere. The atmosphere of such a too tiny moon would boil off in radiation from the nearby star.

The trouble is that moons as big as a terrestrial planet do not seem very easy to build. Astronomers believe that most moons form in much the same way that planets do—gradually coalescing out of a spinning disk of gas, ice and dust [*see box on page 82*].

IN BRIEF

Astronomers are searching for rocky moons that may circle distant exoplanets.
Such exomoons could be a haven for life, provided that the moon is large enough to hold on to an atmosphere.
These moons might be detectable using existing data sets, but their presence would impart such a subtle signal to the data that massive amounts of processing power would be required to find them.

© 2017 Scientific American

Most computer simulations of this piece-by-piece lunar assembly struggle to produce anything much bigger than Jupiter's Ganymede, the largest moon our solar system managed to make. According to the 1997 study, such a moon would need to bulk up fourfold or fivefold to hang on to a permanent atmosphere.

Fortunately, nature has devised other ways to make massive moons. Earth's moon, for example, is too large to have quiescently formed alongside our planet from a shared disk of gas and dust. Many astronomers think, instead, that our Earth-moon system was forged out of a cataclysmic collision early in our solar system's history. Pluto and its largest moon, Charon, are thought to be another collision-forged duo, albeit on a much smaller scale. These pairs could account for other types of moons. In so-called binary-exchange reactions, a giant planet encountering a binary pair captures one member as a moon while the other member gets ejected into space. This kind of exchange has happened at least once before in our solar system: Neptune's biggest moon, Triton, has a bizarre orbit that moves in the opposite direction of the giant planet's rotation. Astronomers believe that Triton is the captured remnant of a binary pair that Neptune tore apart long ago.

These large moons could potentially support liquid water—and thus life—even if they orbit a planet located outside of a star's habitable zone. Extra warmth could come from the reflected light and emitted heat of a host planet, as well as the planet's gravitational pull. Just as the moon raises tides in Earth's oceans, the gravitational tug of a gas giant could send tidal energy rippling through a nearby moon, flexing the lunar interior and pumping it full of frictional heat. The effect is akin to heating up a metal paper clip by bending it back and forth in your hand. Indeed, if a moon orbits too close to its gas-giant planet, it could experience so much tidal heating that it boils off its atmosphere or melts into a glowing ball of slag, according to work by René Heller of the Max Planck Institute for Solar System Research in Göttingen, Germany, and Rory Barnes of the University of Washington. In wider lunar orbits, just the right amount of tidal heating could keep moons comfortably toasty, even if the planet is far from its star's warming rays.

Tidal forces could also change a moon's orbit so that it would eternally present only one hemisphere toward its host planet, just as the moon does to Earth. Envisioning the night skies of such tidally locked worlds, Heller says, yields a deeply strange picture. "Imagine, for example, standing on the planet-facing hemisphere of a tidally locked moon," he says. "The planet would be huge and would not move in the sky. At 'noon' on the moon, which corresponds to the point in its orbit where the star would be highest in the sky, the star would pass behind the planet, and there would be no reflected light from the planet. You would see stars all around but only a black disk directly overhead. At 'midnight,' when the moon's orbit would be taking the star beneath your feet, the planet's illuminated face would shift from a crescent to converge on a full circle, and you'd get all that reflected light. So at midnight, your sky would be much brighter than at noon."

SEARCH STRATEGY

MOONS LARGE ENOUGH to hold on to an atmosphere should, in theory, be visible in data from the Kepler satellite. Since its launch in 2009 until gyroscope problems cut short the mission in 2013, Kepler gazed unceasingly at a single patch of sky, continuously monitoring the brightness of more than 150,000 target stars. It searched for planets by detecting transits: shadows cast toward our solar system as planets crossed the face of their sun. Each transit manifests as a distinct, recurring dip in a star's "light curve," its brightness plotted over time.

The smallest planet Kepler has found in its primary mission, Kepler-37b, is exceedingly small—only slightly larger than Earth's moon. According to Kipping, if Kepler can find moon-size planets, it should also be able to find planet-size moons.

Yet even though Kipping is combing through Kepler's data for signs of them, he is not a member of the Kepler team, nor is his project affiliated with the NASA mission. In fact, anyone could do what he is doing: the Kepler data are publicly available. Astronomers and hobbyists alike have already discovered new planets by sifting through the voluminous data set. Kipping's everyman approach extends to fund-raising as well—he raised $12,000 on a crowdfunding Web site to buy CPUs, then used them to build the Michael Dodds Computing Facility, named for the most generous donor.

Kipping's search strategy is founded on a counterintuitive quirk of gravitational interactions: in a sense, moons orbit planets, but planets also orbit moons. More strictly, a planet and a moon actually orbit a shared center of mass, so that as a moon whips around a planet, the planet wobbles back and forth.

Imagine that you are looking out at a distant moon-planet system. If the moon swings around to the right of the planet, the planet, orbiting the same center of mass, will shift a little bit to the left. Now imagine that moon-planet system transiting left to right across the face of the star. The planet will be left of where it would be if it did not have a lunar companion. This leftward shift, in a left-to-right-moving planet, will delay the onset of the transit by perhaps a few minutes. On the same system's next transit, the moon may be on the other side of its orbit, slightly shifting the position of the planet to the right and advancing the planet's transit a few minutes early.

In addition to these shifts in the onsets of transits, a circling moon can alter the transit's total duration. Unfolding over multiple orbits, this to-and-fro temporal waltz of fluctuating transit properties is an exomoon's expected calling card.

In addition to these timing effects, a sufficiently large moon could block a star's light, adding its own minuscule dip to a transiting planet's signal. The combined planet-moon dip would look much like the signal from an ordinary planet, except for the fact that occasionally the moon would pass directly in front of or behind the planet. The eclipsed moon-planet system would not block quite as much light. Astronomers could use this variation to infer the presence of the hidden moon.

Yet searching for any of these subtle effects has its challenges. A small dip in starlight from a transiting exomoon could just as plausibly be caused by more prosaic phenomena. Every modulation of the light curves so far has been best explained by simple things such as star spots, stellar fluctuations and instrumental errors.

Worse, a single timing signature could be produced by a wide range of possible planet-moon arrangements that varied in details such as the size of the moon and the period and inclination of its orbit. This inherent uncertainty makes it quite difficult to characterize any given exomoon through timing alone.

Yet if astronomers manage to pin down a planet-moon system's orbital configuration through timing effects, as well as the moon's dip in a light curve, they can establish masses for the system's moon, the planet and the star. By pairing those masses with size estimates based on how much starlight a planet or moon

© 2017 Scientific American

How to Make a Moon

Scientists don't expect gaseous Jupiter-like planets to harbor much life, but if such a planet is home to a sufficiently large moon, the moon just might. A fertile moon would have to be massive enough to gravitationally hold on to a thin atmosphere, however. Different methods of moon formation can lead to moons of vastly different sizes.

Lumps from a Disk
Example: Jupiter's moons

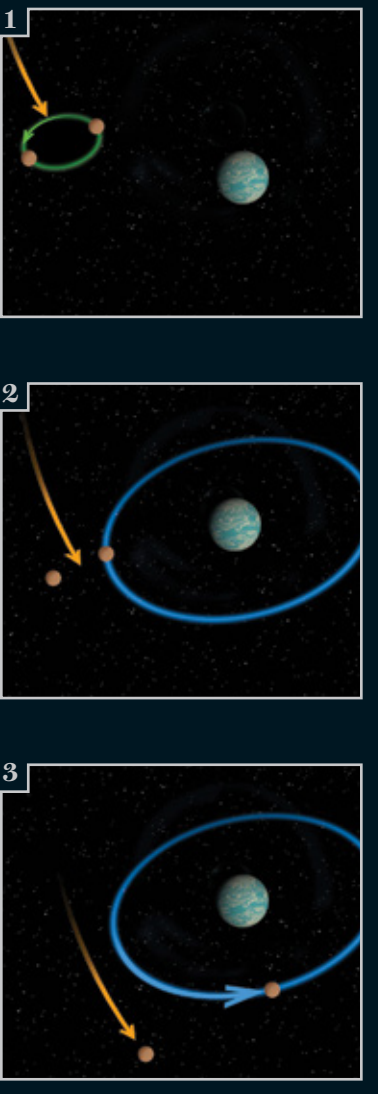
Planets are thought to form out of a disk of dust, gas and ice spinning around a star. Around these young planets additional disks might form, like eddies in a current (*1*). Over millions of years these secondary disks of matter clump into rings and moons (*2* and *3*). Yet these processes can build moons only as big as Jupiter's Ganymede—not large enough to hold on to an atmosphere.

Massive Collision
Example: Earth's moon

Soon after Earth formed, astronomers believe it was struck by a Mars-size object (*1*). The resulting cataclysm spit out a shower of rock and iron (*2*) that, over time, cooled and turned into the moon (*3*). In theory, such collisions could result in two objects that are nearly equal in size. In this double-planet scenario, the "moon" would be just as big as its "planet."

Binary Capture
Example: Neptune's moon Triton

Once a double-planet system forms—perhaps by collision—the pair could encounter another, larger planet (*1*). As the pair flies by (*2*), the larger planet could pull in one of the objects and fling the other off into space (*3*). Captured moons that come from binary planet systems could also be relatively large.



© 2017 Scientific American

blocks, astronomers can infer each object's density, creating a window into the composition, formational history, and potential habitability of planets and their moons. With careful scrutiny of transit after transit for any given system, even more faint details can coalesce from those fluctuations of starlight.

"It's amazing how much can be packed into a light curve," Kipping muses. "It's incredibly satisfying to look up at the stars, these twinkling pinpricks of light in the sky at night, and know that we're able to take this simple measurement of brightness and turn it into all this more complex information." Two years after our conversation, Kipping relocated to Columbia University to continue his exomoon search.

To tease out the presence of a moon orbiting any particular transiting planet, Kipping's HEK project first makes a guess. What would the light curve look like if a moon were orbiting this particular planet? The HEK algorithm generates a very large number of artificial light curves from hypothetical, virtual planet-moon systems that possess a wide variance of masses, radii and orbits. Next it sifts through the Kepler data for matches, gradually homing in on any statistically plausible lunar signals. This exhaustive trial-and-error process is why HEK requires so much computing power. It is also why Kipping prefers to carefully select just the very best targets from Kepler's gargantuan hoard of planets and candidates. Most of those targets are low-mass, Neptune-size worlds that orbit fairly close to a sunlike host star, completing an orbital lap in six months or less. Such planets would manifest the clearest signals of an accompanying large moon. In July of this year Kipping and colleagues found their best-yet candidate for an exomoon, Kepler-1625bi, which could be a Neptune-size moon orbiting a Jupiter-size planet some 4,000 light-years away from Earth. Observations with the Hubble Space Telescope could confirm or refute this candidate in early 2018.

The project also plans to use data from NASA's upcoming Transiting Exoplanet Survey Satellite, launching in 2018, to seek exomoons around red dwarf stars, which are far smaller, dimmer and more numerous than stars like our sun. The small sizes mean that a transiting planet will block a higher percentage of the star's total light. The relatively dim output moves the habitable zone close to the star; any planet orbiting at that radius would have to whip around quickly, giving astronomers more transits to work with. "For us, everything gets better with these stars," Kipping says. "In the very best cases, we could probably detect a moon less than a fifth of an Earth mass."

In perhaps the very worst case, HEK will detect no exomoons at all, a prospect that would at least allow Kipping and his colleagues to set upper limits on how many planets harbor large moons. Already we know something about what is not there. "If there were lots of really big moons, like a two-Earth-radius moon around a Jupiter-size transiting planet, you could just look by eye at the light curve and see the moon's effect," says University of Florida astronomer Eric Ford. "So there's a good chance if that was in the Kepler field, someone would've found it by now or be hot on its trail." After further analysis, Kipping's team has ruled out the possibility that Kepler-22b, one target of the early investigations, has a moon larger than about half the size of Earth.

Other astronomers, such as Eric Agol of the University of Washington, remain skeptical that Kepler's current data set can deliver verifiable exomoons, particularly via temporal effects alone. "My opinion is that a believable detection is going to require actually seeing the transit of a moon," Agol says. "But that's at the very hairy edge of what Kepler can do. Of course, nature can always surprise us."

Despite his doubts, Agol acknowledges that he and a few other collaborators are pursuing an unofficial search of their own, one that, in comparison to HEK, uses less intensive computation to seek more obvious effects in a larger number of Kepler light curves. "My feeling is our search should be around every planet that's been detected, within reason," Agol says.

LUNAR LENSES

KIPPING POINTS OUT that moons can increase the chance for life in more than one way. For example, he says, without the moon, Earth's climate and seasons could be quite different because on astronomical timescales the moon helps to stabilize our planet's tilt. What is more, before the moon spiraled out to its present orbital distance from our world, its enormous tidal effects on the early Earth could have played a vital role in the origin and flourishing of life.

"When we find an Earth-size planet in the habitable zone, one of the first questions will be, 'Well, does it have a moon?'" Kipping says. The answer to that question will help determine whether a planet is a true Earth twin or merely a cousin with a vague family resemblance. "I wonder if our own is a fluke or if things like it are really common," he adds. "With a sample size of one, we can't really know the answer. If we find some outside our solar system, we'll get a better idea."

Through the right kind of telescopic eyes, ones well beyond Kepler's capabilities, an exomoon could do far more than simply signpost a promising mirror Earth orbiting a nearby star. A sufficiently large telescope on the ground or in space could investigate that distant world's atmosphere, looking for markers of life, such as the oxygen that fills our own planet's skies, Kipping says.

Kipping also thinks some exomoons could be used to map the surfaces of their host planets. Astronomers already use transiting planets to map the surfaces of stars by carefully monitoring the star's brightness as the planet crosses its face. "When a moon passes in front of a planet as seen from Earth, you're getting the same opportunity, but now you're looking at the surface brightness of the planet," he explains. "So, potentially, using something very sophisticated, you could begin mapping an Earth twin's continents, its water distribution, all from how the light curve changes shape as the moon passes over. Sometimes I think that's the most likely way we'll ever get anything like a photograph of one of these potentially habitable planets. This could be the first, smallest slice of a very big pie." SA

Lee Billings is an editor covering space and physics at *Scientific American*. His award-winning 2013 book, *Five Billion Years of Solitude,* chronicles the scientific quest to discover Earth-like planets elsewhere in the universe.

MORE TO EXPLORE

The Hunt for Exomoons with Kepler (HEK): I. Description of a New Observational Project. David M. Kipping et al. in *Astrophysical Journal,* Vol. 750, No. 2, pages 115–134; 2012. http://arxiv.org/abs/1201.0752

Five Billion Years of Solitude: The Search for Life among the Stars. Lee Billings. Penguin Group, 2013.

Hunt for Exomoons with Kepler (HEK) project: **www.cfa.harvard.edu/HEK**

scientificamerican.com/magazine/sa

© 2017 Scientific American

An

© 2017 Scientific American

Alien technology is probably not causing Boyajian's star to dim mysteriously. But alternative explanations are hard to come by

By Kimberly Cartier and Jason T. Wright

Illustrations by Victor Mosquera

STRANGE NEWS FROM other Star

© 2017 Scientific American

ONE QUIET AFTERNOON IN THE FALL OF 2014, JUST AS THE TREES WERE changing from green to gold, Tabetha Boyajian visited our astronomy department at Pennsylvania State University to share an unusual discovery. That landscape on the brink of transformation turned out to be a fitting backdrop for a meeting that would change the course of our careers. Then a postdoctoral scholar at Yale University, Boyajian had flagged inexplicable fluctuations of light from a star monitored by NASA's planet-hunting Kepler space telescope. The fluctuations looked nothing like those caused by a planet passing between the star and the telescope. She had already ruled out other culprits, including glitches in Kepler's hardware, and she was looking for new ideas. One of us (Wright) suggested something *very* unorthodox: perhaps the fluctuations in brightness were caused by alien technology.

In the 1960s physicist Freeman Dyson postulated that advanced, energy-hungry civilizations might enshroud their home stars in solar collectors—later called Dyson spheres—to absorb practically all of a star's light. Could this fading star be the first evidence that other cosmic cultures were more than science fiction? That outré idea was a hypothesis of last resort, but for the time being, we could not dismiss it.

The star that stumped Boyajian—now officially known as Boyajian's star and colloquially called Tabby's star—has captivated astronomers and the general public alike. As with all great enigmas, it has generated a seemingly infinite number of possible solutions—none of which wholly explain the curious observations. Whatever is responsible may lie outside the realm of known astronomical phenomena.

UNEARTHED FROM KEPLER'S TREASURE TROVE

BEFORE KEPLER launched in 2009, most planet hunters doggedly revealed new exoplanets (planets orbiting other stars) one by one, like anglers pulling individual fish from the sea. Kepler came along like a trawler, scooping up new worlds thousands at a time.

For four years the telescope continuously observed stars in one small patch of the Milky Way. It was looking for planetary "transits," in which fortuitously aligned worlds cross the face of their host stars and block a fraction of the starlight seen from Earth. Graphed over time, the brightness of a star is described by a so-called light curve. With no transiting planets, the curve will resemble a flat line. Add in a transiting planet, and that light curve will now include U-shaped dips that recur like clockwork each time the orbiting body blocks the star's light. The duration, timing and depth of the dips convey information about the planet itself, such as its size and temperature.

Of the more than 150,000 stars Kepler surveyed, just one—dubbed KIC 8462852, after its number in the Kepler Input Catalog—displayed a light curve that defied explanation. Members of the Planet Hunters citizen science project were the first to notice it when they scoured Kepler's data for transiting worlds overlooked by professional astronomers' automated planet-hunting algorithms. KIC 8462852 showed transitlike dips in starlight seemingly at random, with some lasting a few hours and others persisting for days or weeks. Sometimes the star's light dimmed by about 1 percent (characteristic of the largest transiting exoplanets), but other times it plummeted by up to 20 percent [*see box on page 88*]. No conceivable planetary system could produce such an extreme and variable light curve.

Perplexed, these citizen scientists notified Boyajian, a member of the team overseeing the Planet Hunters project. In 2016 they introduced the star and its mysteries to the world in a peer-reviewed journal article with the subtitle "Where's the Flux?" (Boyajian calls KIC 8462852 the "WTF star").

STRANGE IN MANY WAYS

BOYAJIAN'S STAR held yet more surprises. In the aftermath of the WTF paper, astronomer Bradley Schaefer of Louisiana State University claimed, based on archival data, that Boyajian's star had dimmed by more than 15 percent over the past century.

The claim was controversial because such multidecadal dimming seems next to impossible. Stars stay at nearly the same brightness for billions of years after they are born and only undergo rapid changes just before they die. These "rapid" changes take place on the timescale of millions (rather than billions) of years and are accompanied by clear markers that Boyajian's star lacks. According to every other measurement made, it is an unremarkable middle-aged star. There is no evidence that it is a variable star, pulsing with a regular beat. And there is no indication that it is accreting material from a companion star, no suggestion of anomalous magnetic activity, and no reason to think it might be young and still forming—all phenomena that could rapidly alter its brightness. In fact, aside from its anomalous dimming, this star appears entirely ordinary.

IN BRIEF

KIC 8462852, also known as Boyajian's star, is a strange object discovered more than 1,000 light-years away by the planet-hunting Kepler space telescope.

The star has vexed astronomers with its dramatic and sporadic dimming, a phenomenon difficult to explain via known natural phenomena.

Disks of gas and dust, interstellar debris, comet swarms or even black holes are some of the exotic potential explanations considered by theorists.

Beyond these scenarios is the sensational possibility that the bizarre behavior of Boyajian's star reflects the activities of an advanced cosmic civilization.

© 2017 Scientific American

Yet Schaefer's claim held up when astronomers Benjamin T. Montet and Joshua D. Simon checked it with the original, lesser known Kepler calibration data. They found that Boyajian's star faded by 3 percent over the four-year Kepler mission, an effect as extraordinary as the shorter-term dips.

We must now explain two mind-boggling phenomena related to Boyajian's star: slow dimming over at least four years (and possibly the past century) and deep, irregular dips spanning days or weeks. Although astronomers would prefer a single explanation for both, each phenomenon is difficult to explain on its own and even harder to explain when considered in tandem with the other.

MANY ANSWERS, NONE PERSUASIVE

HERE WE CONSIDER some of the most oft-proposed scenarios to explain the bizarre observations of Boyajian's star. We will judge each on how well it explains the observations and offer our subjective assessment of the probability that the theory is correct.

A DISK OF DUST AND GAS

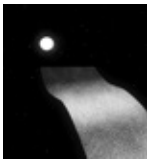
The irregular dips and long-term dimming of Boyajian's star are seen elsewhere—around very young stars with still forming planets. Such stars are belted with circumstellar disks of starlight-warmed gas and dust that, as they form planets, develop clumps, rings and warps. In disks seen edge-on, those features can briefly dim a star's light, and wobbling circumstellar disks can block increasing amounts of a star's light over decades and centuries.

The star is middle-aged, not young, and apparently devoid of a disk. That disk would—as with anything warm—radiate heat as extra infrared radiation, yet Boyajian's star shows no such excess. It could be that the dust and gas exist in a very thin ring sprawled wide around the star, so that the ring blocks starlight along our line of sight without producing much infrared excess. Such rings have never been observed around a middle-aged star such as Boyajian's. Because this scenario can explain Boyajian's star only by invoking a never before seen phenomenon, we judge it to be very unlikely.

A SWARM OF COMETS

Boyajian's original hypothesis was that transiting swarms of giant comets caused the star's dimming. Comets, after all, spend most of their time far from their star and have highly eccentric orbits, which could account for the irregularity of dimming. But what about the lack of heat? The comets would certainly warm up as they approached Boyajian's star and quickly lose that heat as they departed. Any infrared excess would therefore be detectable only during a dip. We detect no infrared excess now, but that absence would make sense if the comets that caused dips a few years ago are now very far from the star, cold and giving off no detectable heat. Even so, any cometary swarm that could also cause the mysterious long-term dimming would have to be very large, unavoidably creating an infrared excess, which, as noted, is missing.

Thus, our verdict is that a cometary explanation is plausible for the dips and very unlikely for the long-term dimming. It stands to reason, though, that if comets are not causing the long-term dimming, then they are probably not causing the dips, either.

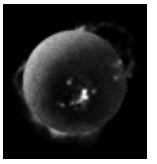


A CLOUD IN THE INTERSTELLAR MEDIUM OR SOLAR SYSTEM

Interstellar space is strewn with gas and dust that diminishes starlight. Maybe an intervening cloud or dense sheet of this material blocks a shifting fraction of light as Kepler's line of sight passes through different parts of it during the telescope's orbit around the sun. Such a cloud could have a density gradient that dims Boyajian's star over long timescales, as well as small knots of material that could cause the extreme short-term dips.

This hypothesis finds some support in the work of Valeri Makarov of the U.S. Naval Observatory and his colleague Alexey Goldin. They argue that some of the smaller dips of light attributed to Boyajian's star are actually deep dips in brightness from fainter adjacent stars in Kepler's field of view, possibly caused by swarms of tiny, dense clouds or comets in interstellar space. We subjectively rank this hypothesis to be plausible.

A related hypothesis suggests that the obscuring cloud may be at the outskirts of our own solar system. In that case, Kepler's orbit around the sun would take the craft's line of sight through such a cloud every year, but we see no annual repetition to the dips of Boyajian's star. Furthermore, there is currently no reason to think that such a cloud exists. Although one could conceive of a cloud of ice and vapor arising from geysers on a Pluto-like body much farther out from our sun, until planetary scientists weigh in on this hypothesis, we rate it conceivable but unlikely.



INTRINSIC STELLAR VARIATIONS

Stars do change in brightness when they begin to exhaust the fuel supply in their core. But this happens on timescales of millions of years, not centuries or days, and at the end, not the middle, of a star's life. Naturally occurring phenomena such as star spots and flares, seen often on our sun, change the brightness of stars on shorter timescales. We may not need to invoke additional orbiting material if the irregular dips and long-term dimming can be explained by brightness variations innate to physical processes within Boyajian's star.

Recently Mohammed Sheikh and his colleagues at the University of Illinois at Urbana-Champaign statistically analyzed the timings, depths and durations of the short-term dips, finding that they are distributed according to a "power law" characteristic of a continuous phase transition (such as magnets realigning themselves in the presence of external magnetic fields). They suggested this distribution could hint that the dips of Boyajian's star are caused by its being on the cusp of an internal transition, such as a global flip of its magnetic field.

But no star like Boyajian's has ever displayed such activity. In fact, the star appears too hot to have the type of stellar dynamo that generates magnetic effects within cooler stars such as our sun. Most problematically, stellar magnetic fields could not create the long-term dimming we see.

Columbia University astronomer Brian Metzger and his colleagues there and at the University of California, Berkeley, have fleshed out a more feasible explanation in which a planet or brown dwarf collided with Boyajian's star. The collision would cause the star to temporarily brighten—and the long-term dimming we see would be the star returning to normal brightness. This scenario does not naturally explain the irregular dips or the

© 2017 Scientific American

detailed shape of the dimming seen by Montet and Simon in the Kepler calibration data, but future studies could solve these problems.

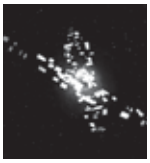
For these reasons, we render a verdict of somewhat plausible for the merger scenario and very unlikely for other explanations invoking intrinsic brightness variations.

BLACK HOLES

One common idea suggested by the public is that a stellar-mass black hole in close orbit around Boyajian's star could block the star's light. That hypothesis fails, however, in three ways. First, because such a black hole would tug the star to and fro in the sky, it would create an easily detectable wobble—a wobble that Boyajian's team looked for and did not detect. Second, stellar-mass black holes are far smaller in size than stars, so one of them would block only a minuscule amount of a star's light. In fact, a black hole's intense gravitational field would counterintuitively magnify most of a background star's light rather than blocking it at all. Third, when a black hole consumes gas and dust, it heats the infalling material so much that it glows brightly at all wavelengths. If there really were a black hole between us and Boyajian's star, we would expect a brightening, not dimming, which we definitely do not see. So, no black hole, right?

Well, not quite. A possible solution involves a distant, free-floating black hole drifting between Boyajian's star and us. Imagine that such a black hole is orbited by a wide, cold disk of material—like the rings of Saturn but larger than our entire solar system—and that this disk possesses an almost transparent outer region and a denser inner region. Such a disk could cause the long-term dimming of Boyajian's star as its nearly invisible outer region, followed by its dense inner region, drifted across our line of sight during the past 100 years. The star's irregular dips might then be shadows cast by rings, gaps and other substructures in the transiting disk. Such a black hole (and its hypothetical disk) would have escaped Boyajian's efforts at high-resolution imaging because it would emit no light itself.

Because we lack observational evidence that black holes host cold, sprawling disks at all, this scenario may seem a bit far-fetched. But theorists predict such disks as a by-product of the supernovae that can give birth to stellar-mass black holes. Moreover, statistical estimates do suggest that such a black hole could have passed in front of at least one of the 150,000 stars monitored by Kepler during the four years of its survey. We subjectively rate this theory as somewhat plausible.



ALIEN MEGASTRUCTURES

Having examined a host of natural explanations for the odd behavior of Boyajian's star and found them lacking, we can now consider the most sensational possibility—an alien megastructure, akin to what Dyson described more than half a century ago.

Imagine that an alien civilization built large numbers of energy-collection panels and that the panels had a range of sizes and orbits around the star. The combined effect of the smaller panels of the swarm would be that they blocked some fraction of the light from the star like a translucent screen.

As denser parts of the swarm come into and out of our line of sight, we might see variations in brightness on scales from hours to centuries. As first noted more than a decade ago by astronomer Luc F. A. Arnold, particularly large panels or groups of panels flying in formation—even bigger, perhaps, than the star itself—would cause huge, discrete dips related to their geometric shape as they transited.

As with the circumstellar disk hypothesis, the lack of infrared emission is a problem. Even alien megastructures must obey fundamental physics, so any energy from starlight they intercept must ultimately be radiated away as heat. This requirement holds no matter how efficient their technology is. Energy cannot be destroyed, so if they are collecting a lot, they must also get rid of a lot in the long run.

There are still ways to make the hypothesis work: a megastructure swarm might radiate its gathered energy away as radio or laser signals instead of heat; it might not form a spherical swarm but a ring precisely aligned with our line of sight; it might use technology beyond our understanding of physics that emits no heat at all. Because of the myriad unknowns, this hypothesis is very hard to test.

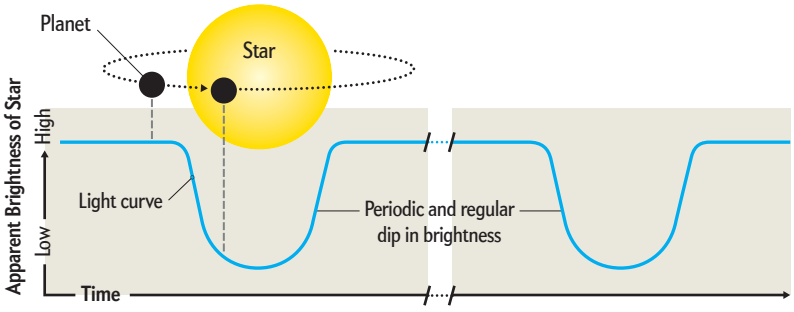
The alien megastructure hypothesis would have to be considered seriously if all natural hypotheses are ruled out. Alternatively, it would find support if we detected obviously artificial radio signals being transmitted from the vicinity of Boyajian's star—a search we have already begun with Boyajian using the Green Bank Telescope in West Virginia. For now, our verdict on

Enigmatic Light Patterns

To astronomers, there is usually no mystery behind a star fading in the sky. Starspots as well as the shadows of planets or debris disks routinely dim the otherwise steady light from mature stars. But none of these explanations seems to apply for one mercurial middle-aged sun known as KIC 8462852—also called Boyajian's star.

Typical Light Curve

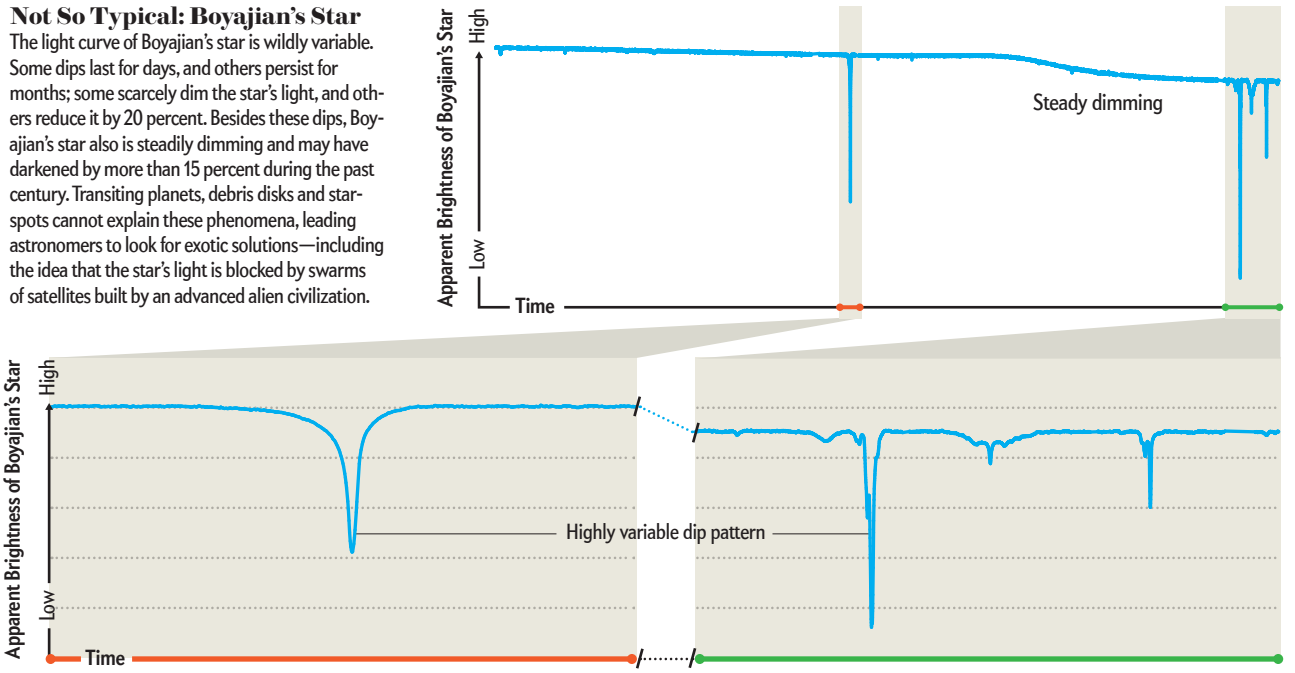
A dimming star can be studied by its light curve—its brightness plotted over time. A planet or disk "transiting" across a star causes a dip in the curve; for planets, this dip recurs every orbital period. Starspots create patterns in light curves based on a star's rotation rate and activity cycle.



© 2017 Scientific American

Not So Typical: Boyajian's Star

The light curve of Boyajian's star is wildly variable. Some dips last for days, and others persist for months; some scarcely dim the star's light, and others reduce it by 20 percent. Besides these dips, Boyajian's star also is steadily dimming and may have darkened by more than 15 percent during the past century. Transiting planets, debris disks and starspots cannot explain these phenomena, leading astronomers to look for exotic solutions—including the idea that the star's light is blocked by swarms of satellites built by an advanced alien civilization.



the most sensational hypothesis to explain Boyajian's star is that its plausibility is unclear: we just do not know enough to put even a qualitative probability on the actions of hypothetical alien life.

AN UNKNOWN, BUT BRIGHT, FUTURE

WHERE DOES this leave us in trying to understand Boyajian's star?

We can rule out any explanation that requires an excess of infrared energy because it is not observed. We can also reject scenarios that require many low-probability events or that invoke physics or objects we have never seen before—at least until all other options have been eliminated.

The best path forward is more fact-finding. Boyajian, now an assistant professor at Louisiana State University, leveraged the public fascination with this star into a successful crowdfunding campaign that purchased time for us on the Las Cumbres Observatory Global Telescope network. On May 20, 2017, this additional monitoring paid off when the star's brightness, once again, started dipping, this time down to almost 3 percent below its usual brightness. Astronomers around the world volunteered labor and telescope time to monitor the star across the electromagnetic spectrum, giving us access to spectral information we lacked during the dips observed by the Kepler mission. We are hopeful these data will help us rule out some of the scenarios here and perhaps generate new hypotheses to explain them.

Other astronomers are seeking out additional archival measurements of the star's brightness to learn more about its long-term dimming. Knowing the timescale of the dimming would further constrain theories about the star's odd light curve and improve our knowledge of how to search for more observational clues.

We also await a better measurement of the distance to Boyajian's star—due from the European Space Agency's Gaia mission—which should help eliminate some hypotheses. If the star is closer than 1,300 light-years, extinction from gas and dust in the interstellar medium cannot explain the current level of dimming. If instead the distance is about 1,500 light-years (the current best estimate), the long-term dimming could perhaps be because of chance patterns of intervening dust along the line of sight. But if the star is much farther away than that, it is far more luminous than previously believed—and then the dimming could be a return to normalcy after a merger, as Metzger's team has suggested.

Unless and until more information trickles in, our speculations about Boyajian's star are limited only by our imagination and a healthy dose of physics. Like the best puzzles in nature, the journey to the truth behind this enigmatic star is far from over. SA

Kimberly Cartier is a senior doctoral student in astronomy and astrophysics at Pennsylvania State University. Her specialty is exoplanets and their parent stars. Cartier also focuses on science communication.

Jason T. Wright is an associate professor of astronomy and astrophysics at the Center for Exoplanets and Habitable Worlds at Penn State and participates in the search for extraterrestrial intelligence. Wright studies exoplanets and the stars they orbit.

SOURCE: "KIC 8462852 FADED THROUGHOUT THE *KEPLER* MISSION," BY BENJAMIN T. MONTET AND JOSHUA D. SIMON, IN *ASTROPHYSICAL JOURNAL LETTERS*, VOL. 830, NO. 2, ARTICLE NO. L39; OCTOBER 20, 2016

MORE TO EXPLORE

Transit Light-Curve Signatures of Artificial Objects. Luc F. A. Arnold in *Astrophysical Journal*, Vol. 627, No. 1, pages 534–539; July 1, 2005.

Planet Hunters IX. KIC 8462852: Where's the Flux? T. S. Boyajian et al. in *Monthly Notices of the Royal Astronomical Society*, Vol. 457, No. 4, pages 3988–4004; April 21, 2016.

KIC 8462852 Faded throughout the *Kepler* Mission. Benjamin T. Montet and Joshua D. Simon in *Astrophysical Journal Letters*, Vol. 830, No. 2, Article No. L39; October 20, 2016.

Avalanche Statistics Identify Intrinsic Stellar Processes Near Criticality in KIC 8462852. Mohammed A. Sheikh et al. in *Physical Review Letters*, Vol. 117, No. 26, Article No. 261101; December 23, 2016.

scientificamerican.com/magazine/sa

Graphic by Tiffany Farrant-Gonzalez

© 2017 Scientific American

Evidence mounts that Saturn's icy moon harbors active hydrothermal vents, making it one of the hottest places to look for life beyond Earth

By Frank Postberg, Gabriel Tobie and Thorsten Dambeck

Illustrations by Ron Miller

UNDER THE SEA OF ENCELADUS

© 2017 Scientific American

© 2017 Scientific American

THE BOTTOM OF THE NORTH ATLANTIC OCEAN MIDWAY BETWEEN BERMUDA and the Canary Islands must rank highly on any list of unlikely places to find a bustling city. Yet there, in the darkness that reigns nearly a kilometer below the sunlit surface, nature has built an undersea metropolis, a complex of limestone towers as tall as skyscrapers that is home to masses of snails, crabs and mussels. The towers form as minerals precipitate out of warm, alkaline water jetting from hydrothermal vents along the ocean floor. Biologists using submersibles and remote cameras found this exotic "Lost City" in the early 2000s and have been studying it ever since to learn how hydrothermal vents can sustain thriving ecosystems so far from the life-giving light of the sun. In the meantime, planetary scientists using the Cassini space probe have made a revolutionary series of related discoveries in the outer solar system, finding strong evidence that hydrothermal vents much like those of Lost City exist not only on Earth but also in the mysterious subsurface ocean of a small, icy Saturnian moon called Enceladus. Could life exist there, too?

Naturally, the possibility of extraterrestrial life tantalizes scientists, but they would have been excited about otherworldly hydrothermal vents even without aliens dangling in front of their eyes. The same evidence that suggests hydrothermal activity on this faraway moon is also providing critical information about the composition and longevity of Enceladus's ocean. Those secrets might otherwise be forever hidden underneath a frozen crust, as they might be for other ocean-bearing moons currently lacking strong evidence for hydrothermal activity, such as Jupiter's Europa.

More fundamentally, the very existence of Enceladus's hydrothermal vents poses an irresistible puzzle. Other than water, the most important ingredient for hydrothermal activity is obviously heat, but this icy moon's sizzling innards are not easily explained. Enceladus is roughly the diameter of England—relatively runty for a moon and far too small to hold onto the primordial heat left over from its formation. Some other source of warmth must be at work deep within. Learning how Enceladus generates and sustains its toasty interior could revolutionize our understanding of icy moons—and their prospects for life.

FIRST HINTS

SCIENTISTS BEGAN TO SUSPECT an ocean within Enceladus in 2005, about a year after Cassini's arrival in the Saturnian system, when the spacecraft observed a huge plume of water vapor and ice grains rising hundreds of kilometers into space from tectonically active terrain around the moon's south pole. In a series of subsequent flybys, Cassini traced the plume to multiple jets emanating from four linear fissures so much warmer than their frigid surroundings that they glowed in infrared. Mission scientists called the fissures "tiger stripes" and pinpointed their jets as the source of the tenuous, sprawling ring of icy particles—known as the E ring—that encloses Saturn's classical ring system. Most of the jets' ice grains, however, travel too slowly to reach this ring and instead fall back to Enceladus as a fine-powdered snow. Based on the 100-meter-high snowdrifts that cover parts of its southern hemisphere, researchers estimate that Enceladus has been venting water into space for 10 million years or more.

Although the "ocean hypothesis" for Enceladus's jets was initially controversial, an extended series of studies by Cassini has now irrefutably confirmed that a global deep sea hides inside the moon. Most recently, an analysis of Enceladus's gravitational field, surface topography and slight rotational wobble by Ondřej Čadek of Charles University in Prague and his collaborators, including one of us (Tobie), placed the best limits yet on the ocean's size and extent. Their work suggests that the crust must be about 35 kilometers thick near Enceladus's equator but less than five kilome-

IN BRIEF

Data from the Cassini orbiter show that Saturn's icy moon Enceladus has a subsurface ocean venting a plume of seawater into space. The plume provides a glimpse of the moon's deep interior.

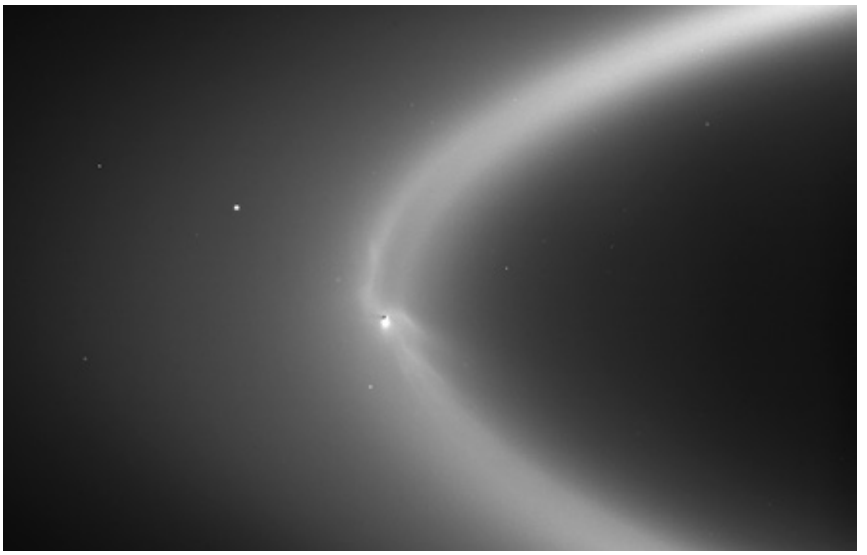
Studies of the moon and its plume have helped researchers estimate the temperature and composition of Enceladus's ocean and have revealed evidence of hydrothermal vents on the seafloor.

On Earth, hydrothermal vents support ecosystems and may have nurtured life's origins. The potential for life on Enceladus depends on the age and longevity of its ocean and hydrothermal activity.

Enceladus's ocean and hydrothermal activity can last only as long as its supply of internal heat. Future missions could clarify this heat source and even make the first discovery of life beyond Earth.

© 2017 Scientific American

ters thick in and around the south pole. Indeed, Alice Le Gall of the University of Versailles Saint-Quentin in France and her collaborators recently detected unusually large thermal microwave emissions at the south polar terrain, indicating an ice shell thickness of only a few kilometers. The ocean floor would lie about 60 to 65 kilometers under the surface, which would mean Enceladus's sea holds about a tenth as much water as the Indian Ocean. And based on data gathered by Cassini in 2009 and 2011, one of us (Postberg) has shown that the water is alkaline and salty, containing sodium chloride—standard table salt—in its ejected plume. This means the ocean most likely lies atop (and leaches minerals from) the moon's rocky core.



ENCELADUS (*center*), seen from two million kilometers away, is embedded in Saturn's E ring, which is formed from the moon's icy plume.

The crucial evidence for the presence of hydrothermal vents was gathered beginning in 2004 by Cassini's Cosmic Dust Analyzer (CDA), before the spacecraft even arrived at Saturn and discovered Enceladus's plume. As Cassini approached Saturn from interplanetary space, unexpected showers of microscopic, fast-moving nanoparticles struck the CDA like buckshot. Years later, after the discovery of the plume, Postberg examined the distribution of the nanoparticle sizes and frequency in the CDA data, finding none of them to be larger than 20 nanometers and all of them to have a composition most consistent with that of essentially pure silicon dioxide—silica, the main constituent of quartz rock and beach sand. Using numerical simulations to trace the most probable trajectories for the silica nanoparticles, Hsiang-Wen Hsu of the University of Colorado Boulder surmised that they had originated in the outer reaches of the E ring. Because we knew that Enceladus produces the E ring, this finding strongly suggested that the nanoparticles had come from the icy moon. Their composition proved to be the smoking gun for discovering Enceladus's hydrothermal activity.

Pure silica jetting from Enceladus was a surprise because its only plausible source would be deep underneath the ice and the ocean, in the moon's rocky core, where silicon mostly exists in minerals chemically bound with other elements such as iron and magnesium. Collisional grinding of those minerals—the rough-and-tumble shattering of rock to make ever smaller pieces—might conceivably create silica nanoparticles. Yet such particles would come in a wide range of sizes, not the very narrow range Cassini observed. Only one other natural explanation remained: the nanoparticles could have crystallized from a supersaturated, silica-rich solution of hot, alkaline water flowing through rock—that is, from hydrothermal vents of exactly the kind found at Lost City on Earth.

A HABITABLE OCEAN?

AT LOST CITY, and presumably on the seafloor of Enceladus, hot water absorbs silica as it flows up through silicate rocks. As the water vents out into the surrounding sea and cools, its capacity to carry absorbed minerals diminishes, and silica nanoparticles form. At this stage, other molecules could glom onto the nanoparticles, making them larger and heavier and thus causing them to eventually precipitate to the bottom—unless, that is, the water were alkaline and not too salty. This relation between the size and longevity of nanoparticles and the temperature and chemistry of their aqueous birthplaces offers researchers an unprecedented window into the environmental conditions of Enceladus's ocean.

Following on Cassini's initial detection of the nanoparticles, a team led by Yasuhito Sekine of the University of Tokyo conducted laboratory experiments to confirm how the nanoparticles formed and to reveal the conditions deep within Enceladus. The scientists found that water at or above 90 degrees Celsius with alkalinity above and salinity slightly below the value of Earth's sea is ideal for creating small, long-lived silica nanoparticles. According to their experiments, the alkalinity of Enceladus's ocean must be between that of terrestrial seawater and ammonia-based household cleaning products. If it were more alkaline than aqueous ammonia, the water's high silica solubility would not allow nanoparticles to form. If the water were less alkaline than seawater on Earth, it would have to be inconceivably hot to dissolve sufficient amounts of silicon dioxide to form silica nanoparticles. Altogether Hsu's, Postberg's and Sekine's respective work raises the possibility that the rich ecosystems of Lost City and other terrestrial hydrothermal vents could conceivably survive and thrive if relocated to the depths of Enceladus. In other words, the ocean of this distant icy moon looks like it could be habitable.

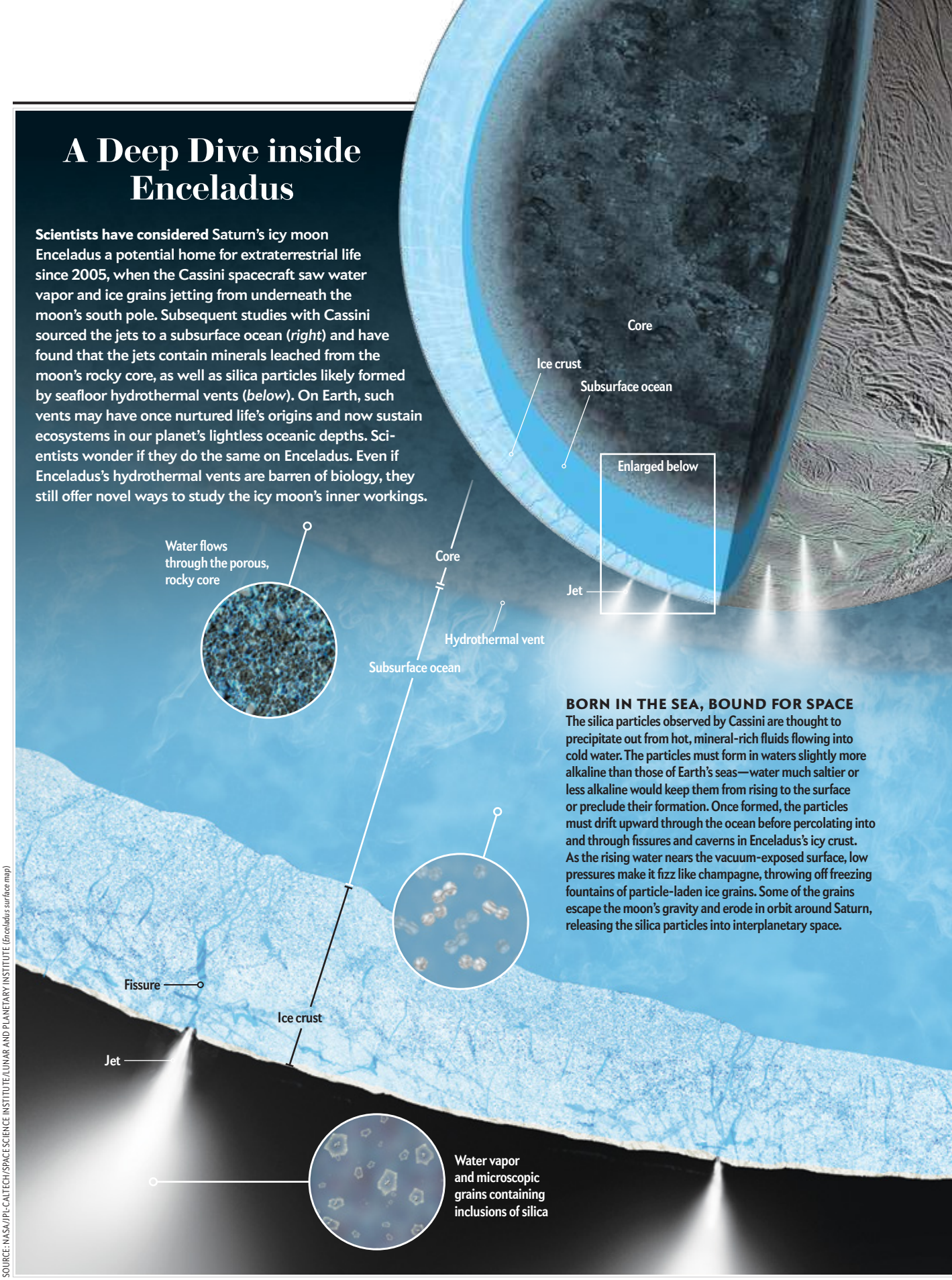
Of course, it could be that Enceladus is currently inhospitable for life and that the silica nanoparticles detected by Cassini are simply relics of ancient hydrothermal activity that ceased long ago. But the work of Sekine and other collaborators suggests this is not the case. In lab experiments and numerical models, freshly formed silica particles average about four nanometers wide, only growing larger over time spans of at least a few months and, at most, a few years. The CDA data show that the typical nanoparticle from Enceladus is between four and 16 nanometers wide, with none wider than 20. Hence, the nanoparticles collected by Cassini must have been created only a short time before being measured. Otherwise, they would have been larger than observed.

COURTESY OF NASA/JPL AND SPACE SCIENCE INSTITUTE

© 2017 Scientific American

A Deep Dive inside Enceladus

Scientists have considered Saturn's icy moon Enceladus a potential home for extraterrestrial life since 2005, when the Cassini spacecraft saw water vapor and ice grains jetting from underneath the moon's south pole. Subsequent studies with Cassini sourced the jets to a subsurface ocean (*right*) and have found that the jets contain minerals leached from the moon's rocky core, as well as silica particles likely formed by seafloor hydrothermal vents (*below*). On Earth, such vents may have once nurtured life's origins and now sustain ecosystems in our planet's lightless oceanic depths. Scientists wonder if they do the same on Enceladus. Even if Enceladus's hydrothermal vents are barren of biology, they still offer novel ways to study the icy moon's inner workings.



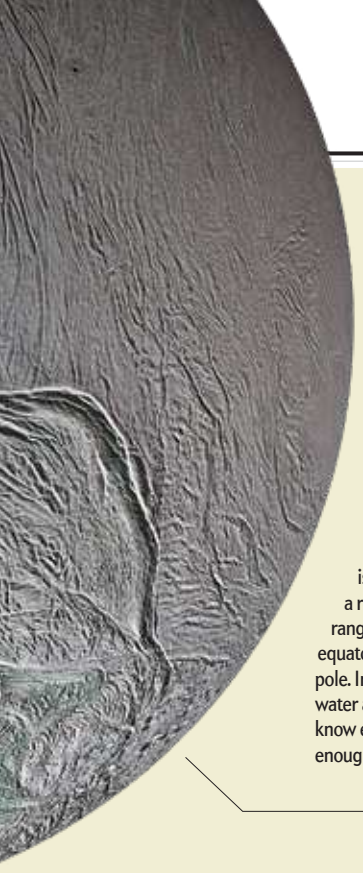
BORN IN THE SEA, BOUND FOR SPACE

The silica particles observed by Cassini are thought to precipitate out from hot, mineral-rich fluids flowing into cold water. The particles must form in waters slightly more alkaline than those of Earth's seas—water much saltier or less alkaline would keep them from rising to the surface or preclude their formation. Once formed, the particles must drift upward through the ocean before percolating into and through fissures and caverns in Enceladus's icy crust. As the rising water nears the vacuum-exposed surface, low pressures make it fizz like champagne, throwing off freezing fountains of particle-laden ice grains. Some of the grains escape the moon's gravity and erode in orbit around Saturn, releasing the silica particles into interplanetary space.

SOURCE: NASA/JPL-CALTECH/SPACE SCIENCE INSTITUTE/LUNAR AND PLANETARY INSTITUTE (*Enceladus's surface map*)

Illustrations by Ron Miller (Enceladus) and Jen Christiansen (heat source icons)

© 2017 Scientific American

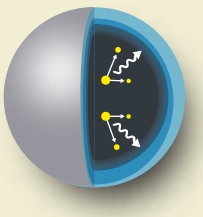


MAPPING A MOON'S INTERIOR

Based on its proposed hydrothermal activity, as well as observations of Enceladus's gravitational field, rotation, surface features and plume composition, researchers now possess a remarkably detailed view of the moon's interior. The subsurface ocean is global in extent, sandwiched between a rocky, porous core and a frozen crust that ranges in thickness from 35 kilometers at the equator to less than five kilometers at the south pole. In total, it probably holds a tenth as much water as the Indian Ocean. Scientists do not yet know exactly how the tiny moon generates enough heat to sustain this large ocean.

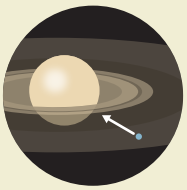
HOW LONG CAN THE OCEAN LAST?

For life as we know it to exist on Enceladus, its ocean must be long-lived and persistent. If instead the ocean is transient, only lasting a few tens of millions of years or cycling through frozen and thawed states, then the moon is likely lifeless. The ocean's longevity is intimately tied to the same mysterious heat supply that powers the moon's hydrothermal activity. Scientists have identified three possible heat sources for the moon, each with different implications for the ocean's longevity.



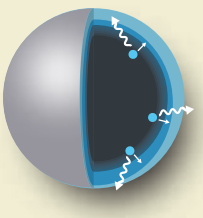
Radioactive Decay

Radioactive isotopes release heat as they decay. Enceladus is thought to be comparable to Earth in isotope abundance, but the moon is so small that it would rapidly deplete and lose this radiogenic heat, causing its ocean to freeze long ago.



Tidal Friction

The gravitational pull of Saturn raises tides within Enceladus that flex its interior, creating heat from friction that could in principle warm the moon for billions of years. Saturn's influence, however, could also alter the moon's orbit in ways that reduce the amount of internal heating, leaving Enceladus cold in only a million years or less.



Serpentinization

"Serpentinization" reactions between water and rock can release heat, but as with radioactive decay, this heat tapers off over time as rock is transformed by the chemical reactions. Yet tidal friction could boost and extend serpentinization by regularly exposing fresh rocks to water percolating through the moon's porous core.

FROM SEAFLOOR TO DEEP SPACE

WE CAN NOW TRACE a typical nanoparticle's journey from the bottom of Enceladus's buried sea out to the wider solar system. After forming at the cooling edges of hot, silica-rich fluids gushing into the cold surrounding ocean, a nanoparticle will spend as much as a few years drifting up through about 60 kilometers of open water.

When it reaches the top of the ocean, the nanoparticle ascends into water-filled fractures that crisscross through the few-kilometer-thick overlying frozen crust of the south polar terrain. Because the seawater is denser than the surrounding ice, its progress upward should halt less than a kilometer underneath Enceladus's surface. But here the so-called champagne effect provides a further boost: as the water, which contains dissolved carbon dioxide and other gases, rises, and the pressure on it decreases, it becomes fizzy with bubbles. The bubbles help to lift the seawater within perhaps 100 meters of Enceladus's surface. Tidal forces may also help the process by repeatedly opening and closing the cracks.

There, we suspect, temporary pools of gas-saturated liquids loaded with suspended particles throw off clouds of ocean mist from bursting bubbles. In such close proximity to the harsh vacuum of space, water evaporates efficiently, although it is near its freezing point. The vapor is dragged into the vacuum and, as if through a chimney, rises up through cracks in the near-surface ice into space. Along with the vapor, the ocean spray from the effervescing pools is carried upward and quickly freezes into micron-sized ice grains, which incorporate the silica nanoparticles like raisins in a bun. Some of the vapor freezes on the walls of ice, releasing latent heat that we see as the infrared glow of Enceladus's surface tiger stripes. The vapor that does not freeze carries the nanoparticle-laden grains up to the surface, hurling them into space as icy fountains.

Most of the grains in the plume fall back to the surface as snow, but those with the highest speeds escape Enceladus to accumulate in the E ring. In the E ring, ionized gas erodes the ice grains and frees the embedded nanoparticles. The liberated nanoparticles then accumulate an electric charge from the ionized gas and free electrons and become the playthings of Saturn's immense electromagnetic fields. Finally, boosted by solar winds, some of the nanoparticles reach velocities up to a million kilometers per hour—about 1 percent of light speed—and zip off into the solar system. A small fraction of the escapees may even reach interstellar space, to surf the voids between the stars.

HOT TOPICS

AS ELABORATE, beautiful and, we believe, true as this narrative is, it does not address what has become the central conundrum of Enceladus: What is the source of the internal heat required to maintain its dynamic ocean? That heat, which is essential for liquid water and life, obviously cannot come from sunlight. The sun's rays are about 99 percent weaker at Enceladus than in Earth's vicinity, giving the icy moon a surface temperature about that of liquid nitrogen.

About half of Earth's internal warmth comes from the slow decay of radioactive isotopes of uranium, thorium and potassium. This radiogenic heating has sustained temperatures exceeding several thousands of degrees C in Earth's interior for billions of years. Although Enceladus probably contains comparable concentrations of radiogenic elements, at only 500 kilometers wide, the diminutive moon loses its internal heat much more efficiently than Earth does.

© 2017 Scientific American

HYDROTHERMAL FLUIDS percolating through Enceladus's hot interior could potentially form seafloor mineral deposits, as in this artist's conception.

In the absence of an additional heat source, Enceladus's interior should be frozen solid. The moon's small size and weak gravity also make its internal dynamics very different from bulky planets like Earth: lower pressures and more moderate temperatures within Enceladus limit the compaction and consolidation of material in its core, allowing water to circulate down through porous rock to create hydrothermal processes in the moon's very center. Contrast this with Earth, where the rapid subterranean increases in pressure and temperature limit water circulation to the top few kilometers of crust.

One might expect the flushing of Enceladus's core to hasten its cooling, sweeping away any radiogenic heat and precluding the high temperatures required to make silica nanoparticles. Yet there is another possible energy source beyond standard radiogenic heating that could explain the moon's present-day hydrothermal activity: tidal heating.

Similar to how Earth's ocean tides arise from our moon and the sun pulling on our planet, tidal heating occurs when a planet's or moon's interior periodically flexes as it moves through a noncircular, eccentric orbit. The flexure from shifting gravitational forces creates friction in a planet's or moon's inner layers, which in turn creates heat. Tidal heating would be particularly powerful within the porous, water-suffused core of a world such as Enceladus. And indeed, data from Cassini clearly show that Saturn's tidal pull profoundly influences the tiny moon—the brightness of its erupted jets and therefore the amount of material ejected vary periodically as the moon whirls around the ringed planet. Evidently, the chimneylike cracks that serve as conduits for mist and water vapor through the ice are at turns squeezed together and pulled apart by tidal tugs that also generate significant amounts of heat.

THE TURNING OF THE TIDE

WHAT WE DO NOT KNOW is whether the ocean we observe today is a transient phenomenon persisting for only tens of millions of years or a sustained feature of the moon that has endured for hundreds of millions or even billions of years. The answer depends on how long the action of tides has heated up Enceladus's interior, which in turn depends on how the moon influences Saturn, as well as its lunar neighbor Dione.

To understand these tidal interactions, we can consider the familiar system of our Earth and moon, which bears some similarities to that of Saturn and Enceladus. Our moon raises tides on Earth, and Enceladus does the same to Saturn. In Earth's ocean, these tidal flows gradually dissipate because of friction against coastlines and the seafloor—an effect that measurably slows Earth's rotation. One hundred years from now, the day will be two milliseconds longer than it currently is, and Earth will have sapped enough of lunar tidal energy to push the moon's orbit out by nearly four meters. Similarly, tidal friction in Saturn's interior infinitesimally affects the giant planet's rotation while increasing Enceladus's distance from Saturn and orbital eccentricity. Higher eccentricities translate into larger tidal effects—and thus more heating—within Enceladus. Early theoretical estimates suggested that Enceladus would generate only weak tidal friction within Saturn, leading to the moon's orbit losing its eccentricity and thus limiting the lifetime of any tidal-heated ocean to no more than one million years.

Recently Valéry Lainey of the Paris Observatory and his collaborators (including Tobie) performed detailed analyses of the motions of Saturn's large moons to place more accurate con-

© 2017 Scientific American

straints on the magnitude of tidal friction in the giant planet's interior. They found that the tidal friction within Saturn is at least 10 times larger than predicted by previous models. If true, this larger figure would mean that Enceladus's orbital eccentricity is stable and long-lived, allowing strong tides that can sustain an ocean for at least tens of millions of years and potentially much, much longer. The longer Enceladus's ocean can persist, it would seem, the greater the likelihood of life's emergence and flourishing there.

BRAVE NEW UNDERWORLDS

IN THE MEANTIME, there is a second possible heat source to consider beyond tidal heating. When water percolates through silicate rock, it can hydrate and change the crystalline structure of certain minerals, releasing substantial amounts of heat in a process called serpentinization. Bolstered by the ready circulation of water through the moon's porous, silicate-rich rocky core, serpentinization could generate several gigawatts of power and be a crucial part of Enceladus's internal heat budget. As long as fresh, unaltered minerals are in contact with circulating water, this heat supply will persist. But as the rock becomes fully serpentinized over millions of years, it will cease producing heat and should cool down in the absence of other influences, such as tidal friction. So it seems that serpentinization alone can hardly sustain a global ocean long enough to allow prebiotic chemistry to evolve.

Even so, serpentinization could still contribute to possible biospheres within Enceladus's depths. On Earth, scientists have observed serpentinization processes powering the hydrothermal vents at Lost City and other undersea sites. Besides making heat, these reactions also produce hydrogen, methane and other organic compounds that sustain the microbes that form the base of the food chain for their isolated, sun-starved ecosystems. Studying such organisms, some researchers have wondered whether life really needs the sun at all.

The idea that similar scenarios are occurring within Enceladus's water-filled core recently got a substantial boost: during the last close flyby on October 28, 2015, Cassini dove deeper into the plume than ever before to allow its ion and neutral mass spectrometer (INMS) to look for hydrogen. In April 2017 the INMS team reported that it indeed detected hydrogen, which makes up about 1 percent of the plume vapor, supporting the idea that serpentinization happens deep in the moon's interior. Serpentinization likely does not contribute substantially to the total heat budget, but the measured ratio of hydrogen and methane in the plume indicated that the so-called methanogenesis reaction could provide a source of chemical energy at hydrothermal sites.

In the late 1980s Michael Russell, then at the University of Strathclyde in Scotland, and his collaborators hypothesized that alkaline hydrothermal vents might have been the birthplace of the first living organisms on the early Earth. Although none were yet known on Earth, Russell argued that such sites would offer a relatively benign yet energy-rich environment in which prebiotic chemistry could brew to form the precursors of modern membranes, metabolisms and self-replicating molecules. Few people took the idea seriously enough to discuss or debate it outside of rarefied academic circles.

The discovery of Lost City catalyzed new interest in Russell's hypothesis, catapulting it to the forefront of contemporary discussions of life's origins. Now the discovery of similar environments within Enceladus—and the potential for their existence in other icy moons such as Jupiter's Europa—is catalyzing another shift in how we think about the possibilities for life elsewhere in the solar system. Biology need not be confined to the warm, wet surfaces of sunlit rocky planets but could perhaps proliferate in a much wider range of environments, sustained in whole or in part by heat from radioisotopes, serpentinization or tidal forces. Enceladus and Europa may be proverbial tips of the iceberg—telltale hints that subsurface oceans also exist in Jupiter's moons Ganymede and Callisto, as well as Saturn's moons Titan and Mimas and even the dwarf planet Pluto. Researchers who, like us, are interested in life beyond Earth are only beginning to grapple with these speculative possibilities and their implications, but it appears increasingly likely that until now, we have drastically underestimated the universe's biological fecundity.

For the time being, we must remain in the dark about whether the interiors of icy moons really do supply all the necessary ingredients for extraterrestrial habitability. The duration and intensity of hydrothermal activity within Enceladus remain an open issue, and discussion of possible hydrothermal activity inside Europa is scarcely more than speculation. Both NASA and its counterpart the European Space Agency are eagerly pursuing answers to these questions and are planning missions to Jupiter's icy moons that could seek Enceladus-style plumes in the late 2020s or early 2030s. In September 2017 Cassini completed its mission—crashing into Saturn to preclude any possibility of contaminating Enceladus or another icy moon with earthly biology. Eventually a new generation of spacecraft could be sent there to undertake in situ investigations, landing on the moon and even gathering samples for return to Earth. At present, such missions exist only in the hopes and dreams of astrobiologists—but perhaps not for long. SA

Frank Postberg directs the evaluation of the mass spectra from Cassini's Cosmic Dust Analyzer. Similarly at home in planetary science, physics and chemistry, he is active both at Heidelberg University and at the University of Stuttgart in Germany, where he studies cosmic dust and icy moons.

Gabriel Tobie is a French planetary scientist who began studying the interiors of icy moons after growing up inspired by NASA's Galileo mission to Jupiter. He now develops models to understand how tidal friction could power the activity of Europa, Enceladus and other icy moons.

Thorsten Dambeck is a German physicist and science writer who focuses on astronomy and planetary science. He first glimpsed Enceladus through a telescope as a youngster in the early 1980s, when Saturn's rings were edge-on, and the tiny moon appeared as a faint white dot.

MORE TO EXPLORE

Tidally-Induced Melting Events as the Origin of South-Pole Activity on Enceladus. Marie Běhounková et al. in *Icarus,* Vol. 219, No. 2, pages 655–664; June 2012.

Ongoing Hydrothermal Activities within Enceladus. Hsiang-Wen Hsu et al. in *Nature,* Vol. 519, pages 207–210; March 12, 2015.

Keeping Enceladus Warm. B. J. Travis and G. Schubert in *Icarus,* Vol. 250, pages 32–42; April 2015.

Enceladus's Internal Ocean and Ice Shell Constrained from Cassini Gravity, Shape, and Libration Data. Ondřej Čadek et al. in *Geophysical Research Letters,* Vol. 43, No. 11, pages 5653–5660; June 16, 2016.

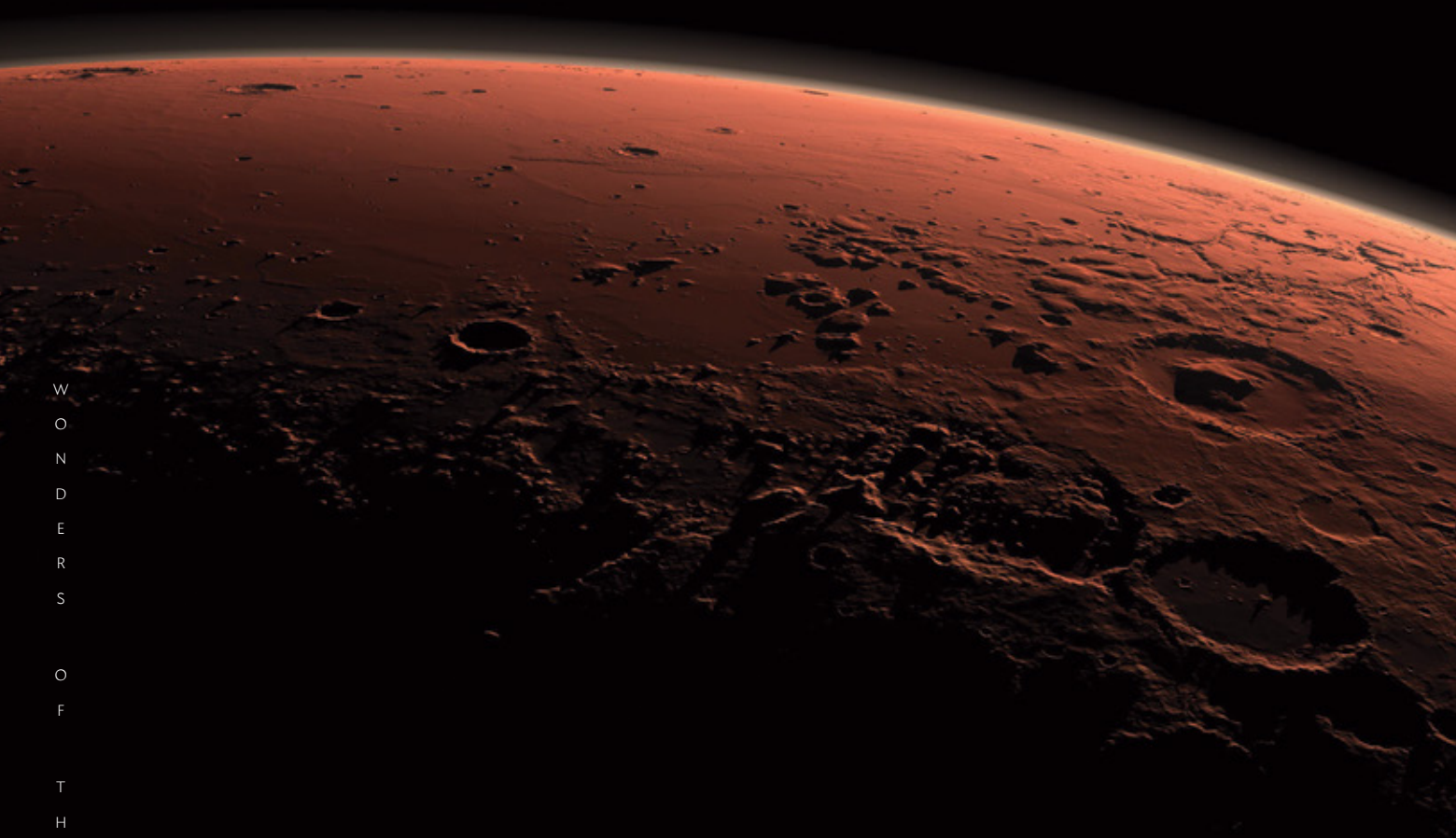
Thermally Anomalous Features in the Subsurface of Enceladus's South Polar Terrain. Alice Le Gall et al. in *Nature Astronomy,* Vol. 1, Article 0063; March 13, 2017.

Cassini Finds Molecular Hydrogen in the Enceladus Plume: Evidence for Hydrothermal Processes. J. Hunter Waite et al. in *Science,* Vol. 356, pages 155–159; April 14, 2017.

scientificamerican.com/magazine/sa

© 2017 Scientific American

Experiments now under development could finally answer one of the



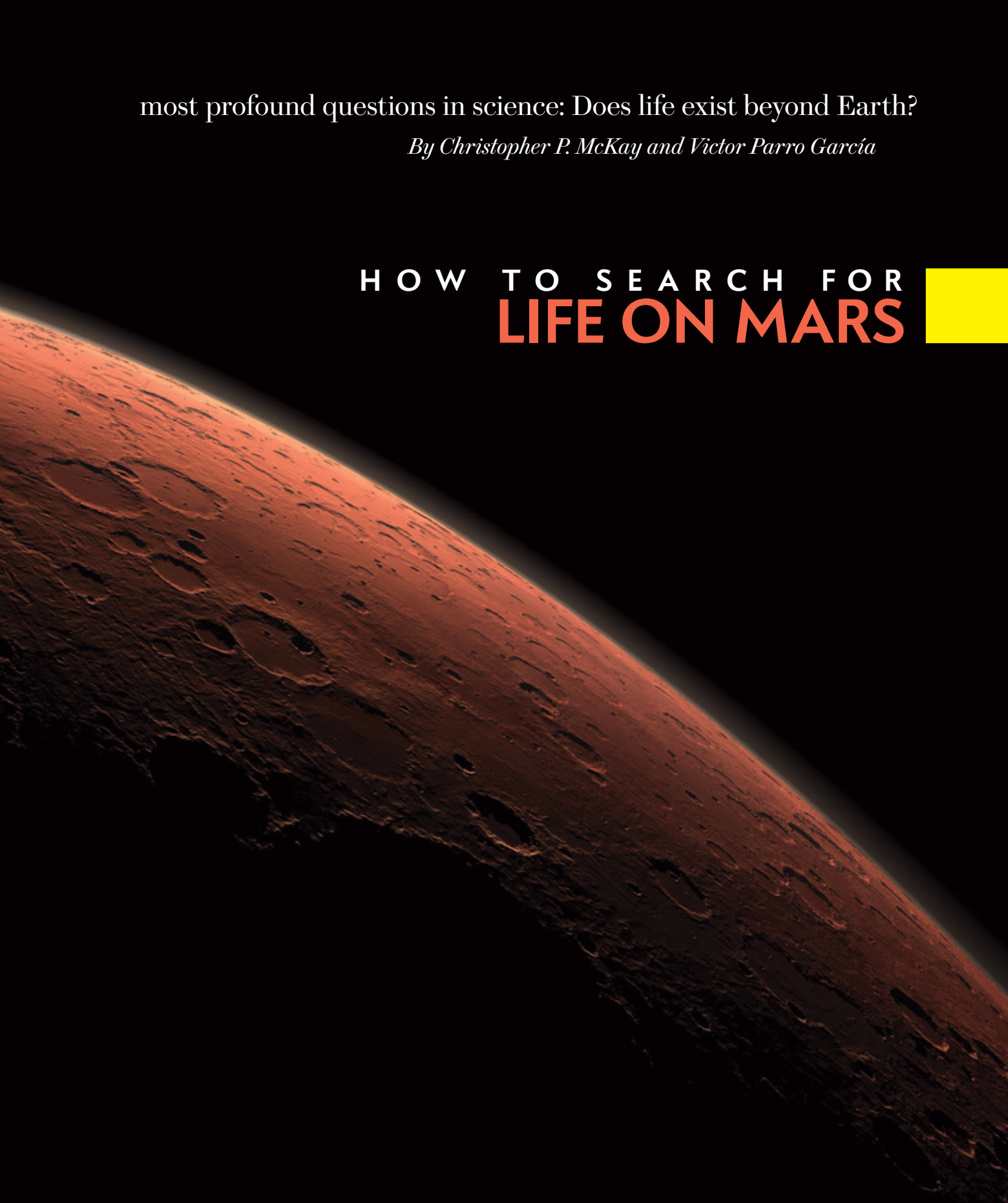
GALE CRATER—seen to the left of center in this composite image, distinguished by the mountain inside—once held liquid water. In 2014 the Curiosity rover found evidence of organic molecules on its plains.

© 2017 Scientific American

most profound questions in science: Does life exist beyond Earth?

By Christopher P. McKay and Victor Parro García

HOW TO SEARCH FOR LIFE ON MARS



© 2017 Scientific American

Astronomers have learned a great many things about Mars since the first probes landed there nearly four decades ago.

We know that liquid water once flowed across its surface and that Mars and Earth were similar in their early history. When life on Earth arose, some 3.5 billion years ago, Mars was warmer than it is today and had liquid oceans, an active magnetic field and a thicker atmosphere. Given the similarity between the two planets, it seems reasonable to think that whatever steps led to life on Earth could also have occurred on Mars.

In fact, for all we know, microscopic life might still exist on the Red Planet. Every mission to our neighbor in the past 35 years has examined its geology, not its biology. Only the twin Viking 1 and 2 spacecraft, which touched down in 1976, conducted the first and thus far only search for life on another world. Each spacecraft carried four experiments relevant to the search, and each experiment returned ambiguous data. The Viking missions gave us puzzles, not answers. Yet we now know that Viking's methods would not have been able to find life on Mars even if it were there—which means the question of whether the planet harbors life remains open.

Fortunately, in the intervening decades microbiologists have developed a cornucopia of tools for detecting microorganisms. These tools are now unexceptional here on Earth. But if applied by one of the several missions soon expected to head for Mars, they could deliver a first: a definitive answer as to whether our closest neighbor also pulses with life.

THE FIRST SEARCH

THE VIKING EXPERIMENTS looked for life using standard search techniques of the time. In the initial experiment, the lander took a scoop of Martian soil and added carbon compounds as food for any microorganisms that might be in the soil. If microbes did exist in the soil, we might expect them to consume the food and release carbon dioxide.

In fact, the Viking missions did see this behavior. On its own, the test would seem to indicate that microorganisms were present in the Martian soil. When combined with the results of the other experiments, however, researchers could not be sure.

The second experiment looked for evidence of photosynthesis but returned inconclusive results. A third experiment added water into a soil sample. If life were present, the moist soil might have produced carbon dioxide. Instead it produced oxygen. This was very strange, as no known soil on Earth does this. Scientists concluded that the oxygen came from a chemical reaction.

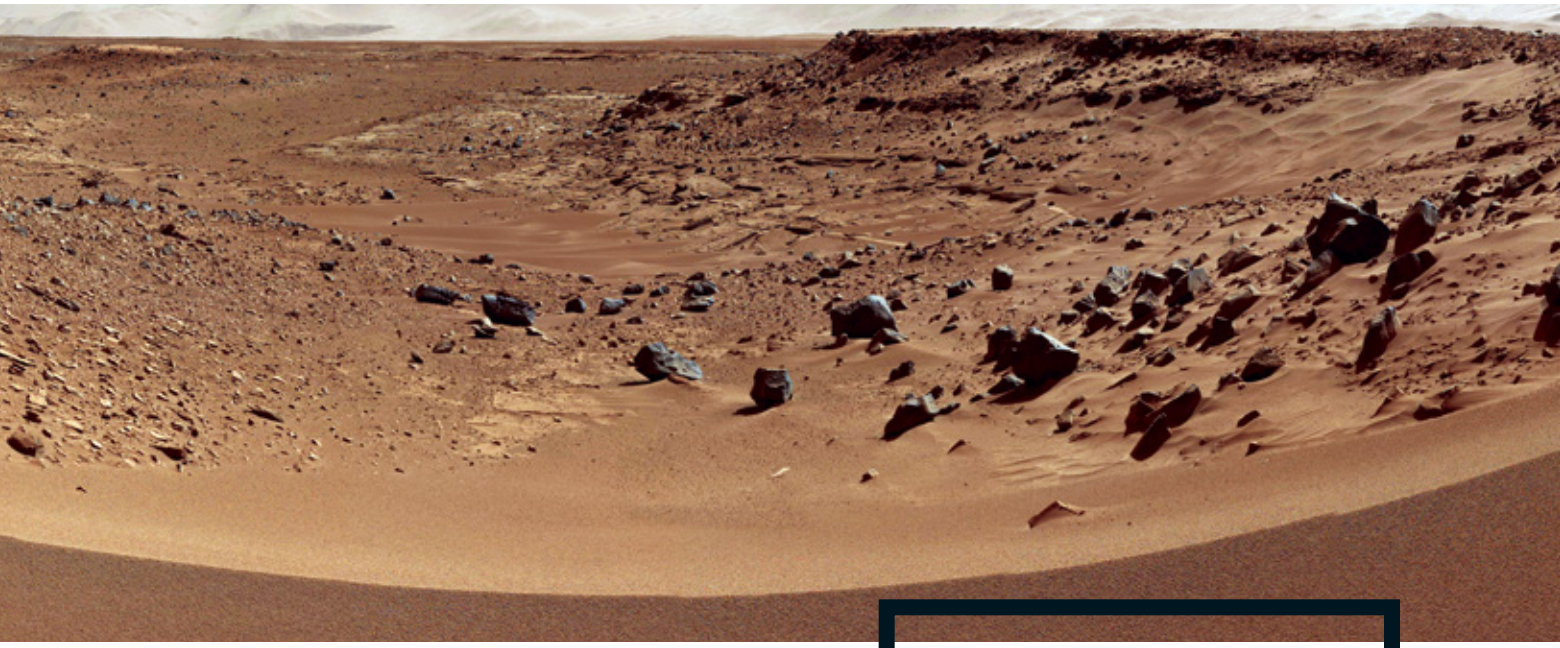
In the final experiment, the landers searched for organic compounds in the soil. Organics are carbon-containing compounds that form the building blocks of life. If any life existed on Mars, we would expect to find these compounds. Yet organics alone would not provide definitive evidence of life, because we also expect meteorites to continuously deposit organic com-

IN BRIEF

No mission to Mars has searched for life since the Viking program in the 1970s. Those missions did not find convincing evidence for life, and we now know that their experiments were doomed to fail. **A modern search for life** on Mars could employ biological tests that we commonly use on Earth. **Such experiments** could be included on a number of missions that are scheduled to travel to Mars by the end of this decade.

PRECEDING PAGES: COURTESY OF NASA, JPL/CALTECH;
THIS PAGE AND OPPOSITE PAGE: COURTESY OF NASA, JPL/CALTECH AND MALIN SPACE SCIENCE SYSTEMS

© 2017 Scientific American



DRY NOW, Gale Crater once held a freshwater lake that could have supported life. The Curiosity rover's successor will be able to carry out experiments that could determine whether traces of life still exist.

pounds on Mars. Puzzlingly, the experiment found no evidence of organics whatsoever.

Taken together, the findings left investigators stumped. Most scientists believed that chemical reactions were responsible for the results in the last two experiments, but chemistry could not quite explain the first. A small but vocal minority of Mars scientists held that the first experiment did, in fact, find evidence of life. But most everyone else concluded that Mars was barren.

In 2008, 32 years after Viking landed, the solution to these puzzles began to emerge when NASA's Phoenix lander touched down in the northern polar region of Mars. To everyone's surprise, Phoenix detected perchlorate, a rare molecule on Earth that features four oxygen atoms connected to a chlorine ion, which are connected to a magnesium or calcium ion. When perchlorate salts reach 350 degrees Celsius, they decompose, releasing reactive oxygen and chlorine. Perchlorates are so reactive that they are used in many rocket fuels.

This finding made investigators see that the perchlorates could well have obliterated signs of life in the soil. Viking's organic search experiment first heated the soil sample to 500 degrees C so that it might vaporize any organic molecules and detect them in gaseous form. But in 2010 a team led by Rafael Navarro-González of the National Autonomous University of Mexico, which included one of us (McKay), showed that perchlorate would have completely destroyed any carbon compounds in the soil during the heating process.

Perchlorate also illuminates the puzzles of the first and third experiments. In the first experiment, adding food to the soil generated carbon dioxide. But perchlorate produces bleachlike compounds when exposed to cosmic rays. These compounds can decompose organic molecules (such as those found in the added food), producing carbon dioxide in the process. In the third experiment, oxygen emerged from moistened soil. The perchlorate bleach production also forms oxygen, yet the oxygen remains initially trapped in the soil. It is released only later, once the soil is wetted, as happened on Viking. Two mysteries solved.

Yet hope for the discovery of life is not lost. The Curiosity rover landed on Mars in 2012 and has been taking samples of the soil ever since. In early 2014 the Sample Analysis at Mars (SAM) instrument team (which includes McKay), led by Paul Mahaffy of NASA's Goddard Space Flight Center, reported that the experiment found carbon compounds in ancient mudstone sediments on the bottom of Gale Crater, even in the presence of perchlorate. Later, in 2015, the SAM team reported the presence of chlorobenzene in one of the Martian samples. Complex organics exist on Mars—Viking was just unable to find them. Could the same be true for life itself?

MODERN APPROACHES

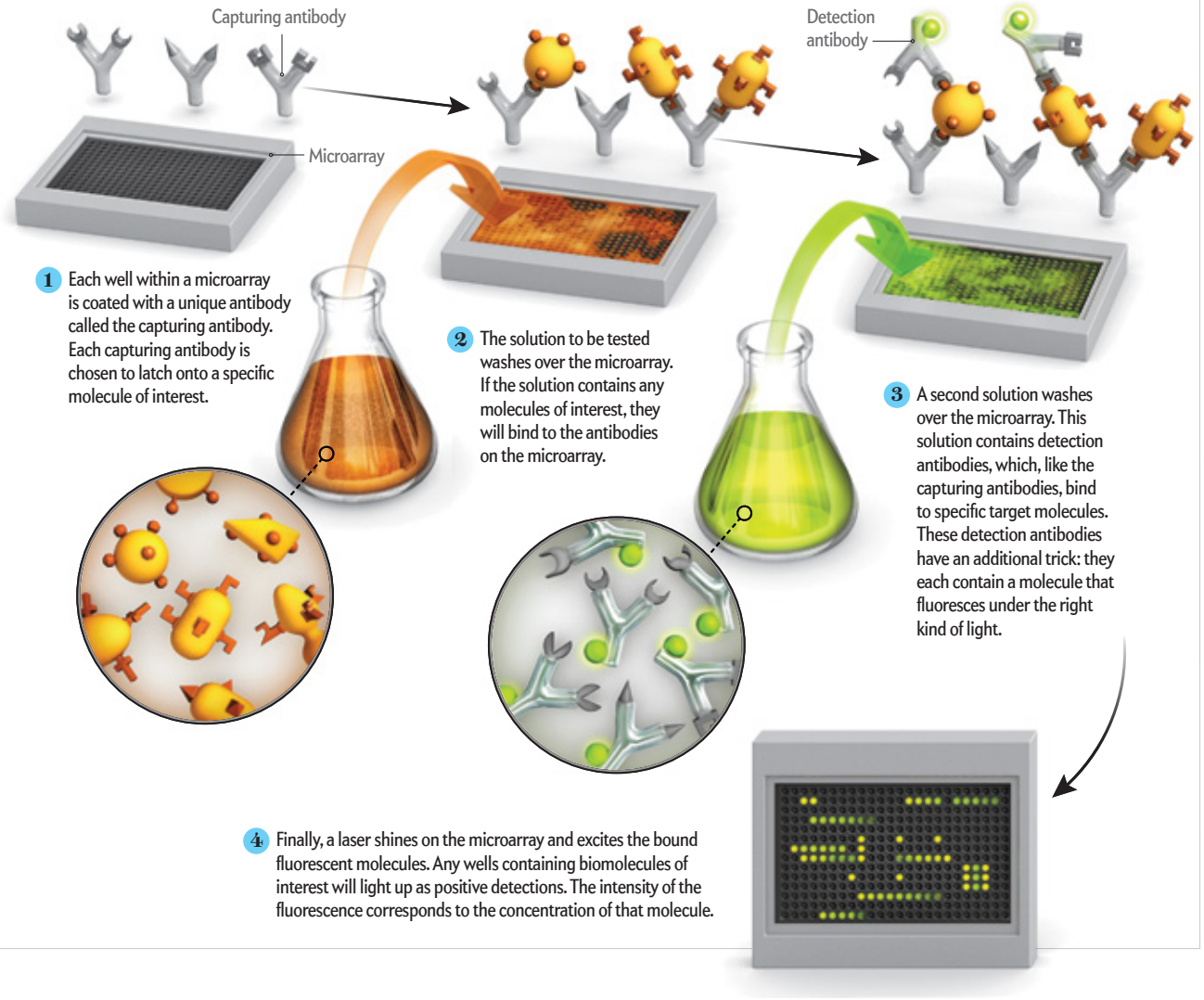
IN THE 40 YEARS since the Viking landers were built, microbiology technology has changed dramatically. The Viking missions used culture-based methods, in which microorganisms grow in petri dishes. But these are no longer considered definitive, and we now know that only a small fraction of soil microbes can be cultured. Scientists have developed vastly more sensitive techniques that directly detect the biomolecules in microbial lifeforms. These new methods provide the basis for a novel way to search for evidence of life on Mars.

The most widely known method is DNA detection and sequencing. No longer is it necessary to culture an organism so that it will replicate sufficiently to provide enough DNA for sequencing. Several teams are working on ways to incorporate DNA-extraction technologies into instruments suitable for upcoming Mars missions.

© 2017 Scientific American

The Martian Protein Detector

Immunoassay tests exploit the Velcro-like nature of antibodies—Y-shaped proteins found in the immune system—to act as precise detectors of foreign molecules. A single immunoassay test can detect hundreds of biological molecules (such as proteins), as well as molecule fragments. These tests are being optimized to search for evidence of life on Mars.


Capturing antibody
Microarray
Detection antibody
1 Each well within a microarray is coated with a unique antibody called the capturing antibody. Each capturing antibody is chosen to latch onto a specific molecule of interest.
2 The solution to be tested washes over the microarray. If the solution contains any molecules of interest, they will bind to the antibodies on the microarray.
3 A second solution washes over the microarray. This solution contains detection antibodies, which, like the capturing antibodies, bind to specific target molecules. These detection antibodies have an additional trick: they each contain a molecule that fluoresces under the right kind of light.
4 Finally, a laser shines on the microarray and excites the bound fluorescent molecules. Any wells containing biomolecules of interest will light up as positive detections. The intensity of the fluorescence corresponds to the concentration of that molecule.

One drawback of relying on DNA detection to reveal life on Mars is that although DNA is common to all life on Earth, it may not occur in alien life. Or if it is present, it may be so different that DNA detectors built to find Earth biology will miss it.

Fortunately, Mars could harbor other signs of life. Among these biomarkers are proteins and polysaccharides. Proteins are long, linear chains composed of various mixtures of the 20 different kinds of amino acids used by life. Amino acids are present in meteorites and are likely to have been a common component of the prebiotic environment on any world. Polysaccharides are long chains of sugars constructed by enzymes (biological catalysts), which themselves are proteins.

Detecting molecules as complex as a protein or a polysaccharide would be strong evidence of life, widely defined: a biological system that encodes information and uses this information to build complex molecules. These complex molecules would stand out against any background of simple prebiotic molecules like a skyscraper would stand out against a field of boulders.

One of us (Parro García) has been developing an instrument for detecting such complex molecules on Mars. It is based on a technique—immunoassay testing—already in use for simultaneously detecting hundreds of different types of proteins, polysaccharides and other biomolecules (including DNA itself).

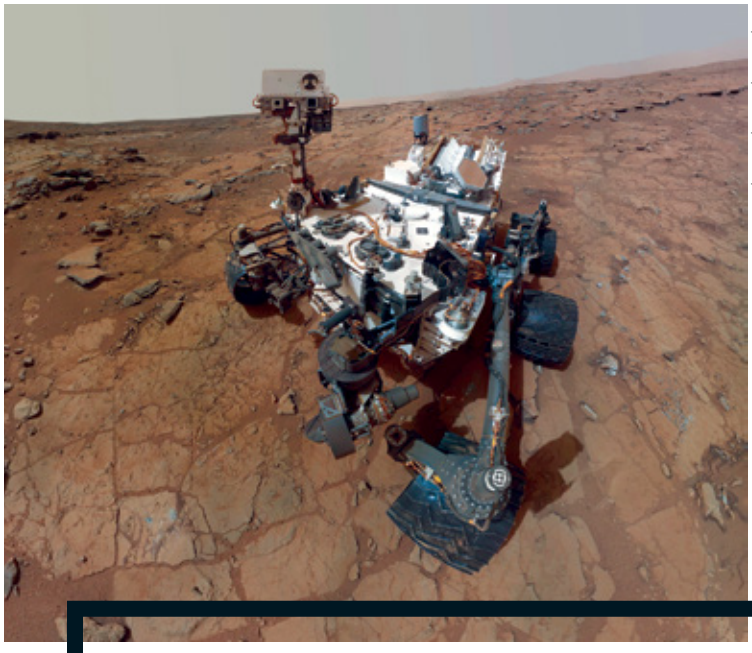
Immunoassay tests employ antibodies—Y-shaped proteins—each of which binds to just one type of biomolecule [*see box above*]. In an immunoassay test, a solution that might contain substances of interest is poured over a large array of antibodies, each one designed to bind to a specific target. If the sample solution contains a biomolecule that links to an antibody in the array, the antibody will capture and, by binding, identify it.

Illustration by Emily Cooper

© 2017 Scientific American

One nice feature of immunoassays is that antibodies can detect molecules that are smaller and less complex than full proteins are. The test can thus search for molecules that are life-related but of lesser complexity, such as fragments of proteins that have broken into bits. Finding these bits would also imply that life exists.

All of Earth's organisms collectively contain many millions of different proteins. With so many to choose from, how do we pick the few hundred that a single immunoassay test could search for? The short answer is that we cannot know for sure. But we can make educated guesses based on two strategies: First, we could search for proteins that would be useful or essential to survival on Mars. For example, we might search for enzymes that consume perchlorate, cold-adapted enzymes that would allow a microorganism to survive Mars's frigid temperatures or enzymes that would repair damage to DNA caused by Mars's strong ionizing radiation. Second, we could target molecules that are ubiquitous throughout the microbial world, such as peptidoglycan, which is a universal component of all bacterial cell walls, or adenosine triphosphate (ATP), which is used by all living organisms on Earth to transport chemical energy for metabolic activity.



CURIOSITY SELF-PORTRAIT shows where it first drilled into the soil. Any microbes would have to live far below the surface.

Even if Mars's harsh environment has destroyed large molecules such as DNA and proteins, we might still find evidence for life in the debris. Many types of molecules are chemically equivalent to one another but may have opposite "chirality"—their bonds may twist to the left or the right. Life on Earth is dominated by left-handed amino acids. If an experiment detects amino acids and finds a particular set that has a dominant left- or right-handed chirality, this would be compelling evidence for the presence of life. Interestingly, if that chirality were right-handed—the opposite of Earth proteins—it would be evidence that the life-forms on Mars evolved independently from life on Earth.

MISSION PLANNING

VIKING CARRIED THREE biology experiments; we might imagine a mission to Mars that also carries three biomarker search instruments—perhaps a DNA detector, an immunoassay microchip, and an instrument to detect and characterize amino acids. The technology is nearly ready. The next task is to pick a target—the location that holds the best chance of harboring biomarkers.

Ice and salt are friends to biomarkers, protecting them from damage and decay. The enemies? Ionizing radiation and heat. Fortunately, the low temperatures on Mars make thermal decay negligible even over the age of the planet. Ionizing radiation, however, could completely destroy biomarkers that are within the first meter or so of the surface over a few billion years. The promising targets, then, are icy sites that may have harbored recent life—such as the Phoenix landing site near Mars's north pole—or sites where erosion has recently exposed the ancient material. In either case, one would want to drill down to extract samples from a meter or more below the surface.

The missions to Mars that are now being planned could conduct this search. The European ExoMars mission, slated for 2020, should be able to carry a drill. NASA is working on a mission that will launch another copy of the Curiosity rover in 2020. ExoMars and the new Curiosity could search the dry equatorial regions of Mars for biomarkers in salt and sedimentary deposits. (Neither rover can function in the polar regions.)

As for a polar search, NASA is studying an inexpensive lander called Icebreaker that could do the job. Equipped with a one-meter drill and an immunoassay instrument, it could search the water-rich northern permafrost of Mars for biomarkers in the ice-cemented ground.

Any one of these missions would be a worthy candidate to lead the next era of Mars exploration. The past few decades of research have left no doubt that Mars once harbored liquid water. The time has now come to test whether that once watery planet provided a home to any life-forms. If we find biological molecules on Mars—and especially if those molecules indicate that Martian life arose independently of Earth life—we will gain a profound insight into life beyond our home. Just as we know that there are many stars and many planets, we will know that there are many biologies. We will know that the universe is alive with diversity. SA

Christopher P. McKay is a scientist at NASA's Ames Research Center.
Victor Parro García is a scientist at the Center for Astrobiology in Spain.

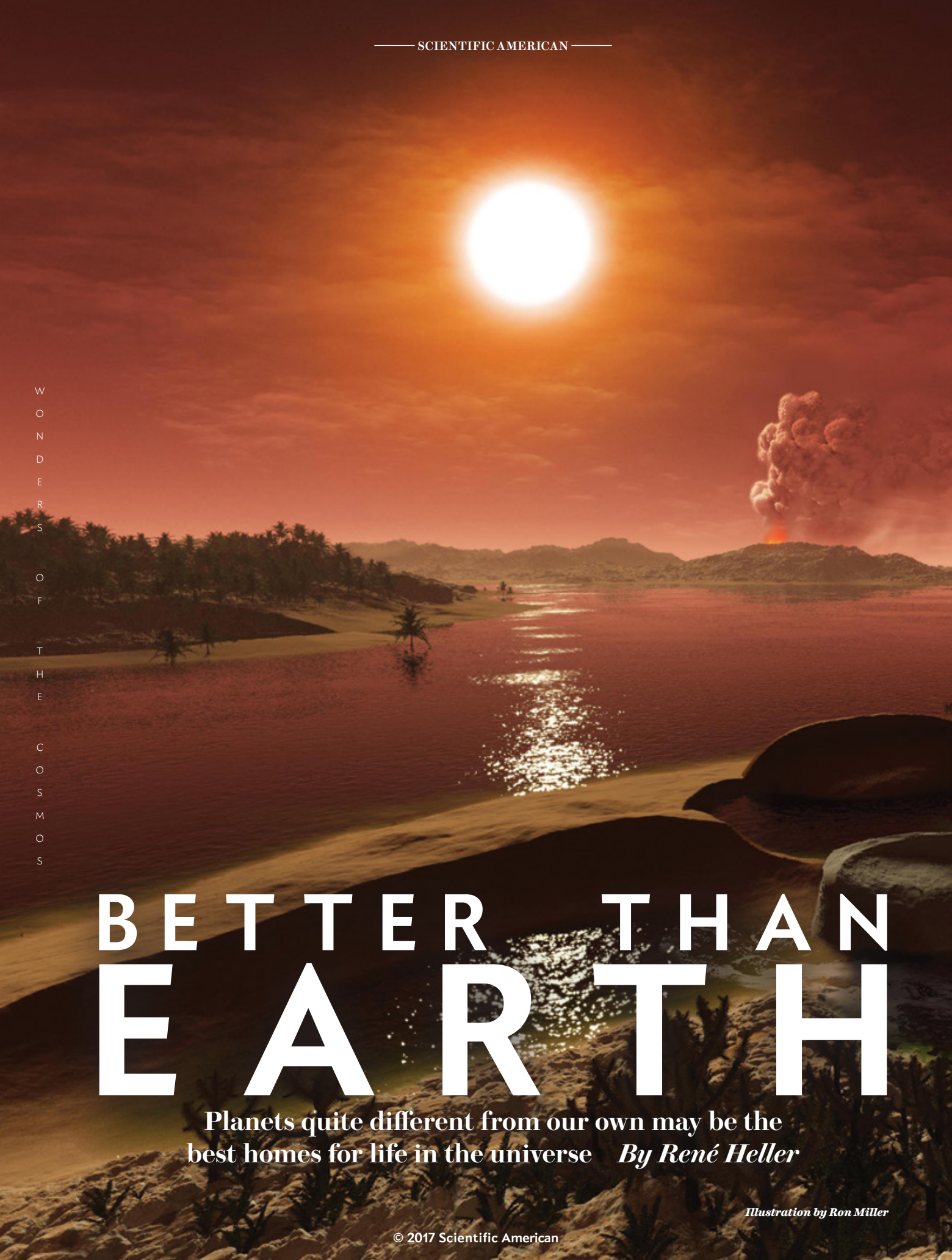
MORE TO EXPLORE

Signs of Life Detector (SOLID) experiment: http://auditore.cab.inta-csic.es/solid/en
NASA's Mars Exploration Program: http://mars.jpl.NASA.gov

scientificamerican.com/magazine/sa

COURTESY OF NASA, JPL/CALTECH AND MALIN SPACE SCIENCE SYSTEMS

© 2017 Scientific American



BETTER THAN EARTH

Planets quite different from our own may be the best homes for life in the universe *By René Heller*

Illustration by Ron Miller

© 2017 Scientific American

LARGE, ROCKY "superhabitable" world orbiting a star smaller than our sun might be both familiar and alien. The landscape would be flattened by higher surface gravity, and plants there could be tinted darker than Earth's greenery to better absorb the faint starlight.

© 2017 Scientific American

DO WE INHABIT THE BEST OF ALL POSSIBLE WORLDS? GERMAN MATHEMATICIAN Gottfried Leibniz thought so, writing in 1710 that our planet, warts and all, must be the most optimal one imaginable. Leibniz's idea was roundly scorned as unscientific wishful thinking, most notably by French author Voltaire in his magnum opus, *Candide*. Yet Leibniz might find sympathy from at least one group of scientists—the astronomers who have for decades treated Earth as a golden standard as they search for worlds beyond our own solar system. Because earthlings still know of just one living world—our own—it makes some sense to use Earth as a template in the search for life elsewhere, such as in the most Earth-like regions of Mars or Jupiter's watery moon Europa. Now, however, discoveries of potentially habitable planets orbiting stars other than our sun—exoplanets, that is—are challenging that geocentric approach.

Over the past two decades astronomers have found more than 1,800 exoplanets, and statistics suggest that our galaxy harbors at least 100 billion more. Of the worlds found to date, few closely resemble Earth. Instead they exhibit a truly enormous diversity, varying immensely in their orbits, sizes and compositions and circling a wide variety of stars, including ones significantly smaller and fainter than our sun. Diverse features of these exoplanets suggest to me and to others that Earth may not be anywhere close to the pinnacle of habitability. In fact, some exoplanets, quite different from our own, could have much higher chances of forming and maintaining stable biospheres. These "superhabitable worlds" may be the optimal targets in the search for extraterrestrial, extrasolar life.

AN IMPERFECT PLANET

OF COURSE, our planet does possess a number of properties that, at first glance, seem ideal for life. Earth revolves around a sedate, middle-aged star that has shone steadily for billions of years, giving life plenty of time to arise and evolve. It has oceans of life-giving water, largely because it orbits within the sun's "habitable zone," a slender region where our star's light is neither too intense nor too weak. Inward of the zone, a planet's water would boil into steam; outward of the area, it would freeze into ice. Earth also has a life-friendly size: big enough to hold on to a substantial atmosphere with its gravitational field but small enough to ensure gravity does not pull a smothering, opaque shroud of gas over the planet. Earth's size and its rocky

IN BRIEF

Astronomers are searching for twins of Earth orbiting other sunlike stars. **Detecting Earth-like twins** remains at the edge of our technical capabilities. **Larger "super-Earths"** orbiting smaller stars are easier to detect and may be the most common type of planet. **New thinking suggests** that these systems, along with massive moons orbiting gas-giant planets, may also be superhabitable—more conducive to life than our own familiar planet.

© 2017 Scientific American

composition also give rise to other boosters of habitability, such as climate-regulating plate tectonics and a magnetic field that protects the biosphere from harmful cosmic radiation.

Yet the more closely we scientists study our own planet's habitability, the less ideal our world appears to be. These days habitability varies widely across Earth, so that large portions of its surface are relatively devoid of life—think of arid deserts, the nutrient-poor open ocean and frigid polar regions. Earth's habitability also varies over time. Consider, for instance, that during much of the Carboniferous period, from roughly 350 million to 300 million years ago, the planet's atmosphere was warmer, wetter and far more oxygen-rich than it is now. Crustaceans, fish and reef-building corals flourished in the seas, great forests blanketed the continents, and insects and other terrestrial creatures grew to gigantic sizes. The Carboniferous Earth may have supported significantly more biomass than our present-day planet, meaning that Earth today could be considered less habitable than it was at times in its ancient past.

EARTH IS PAST ITS PRIME, AND THE BIOSPHERE IS NEARING ITS END. ALL THINGS CONSIDERED, OUR PLANET IS ONLY MARGINALLY HABITABLE.

Further, we know that Earth will become far less life-friendly in the future. About five billion years from now, our sun will have largely exhausted its hydrogen fuel and begun fusing more energetic helium in its core, causing it to swell to become a "red giant" star that will scorch Earth to a cinder. Long before that, however, life on Earth should already have come to an end. As the sun burns through its hydrogen, the temperature at its core will gradually rise, causing our star's total luminosity to slowly increase, brightening by about 10 percent every billion years. Such change means that the sun's habitable zone is not static but dynamic, so that over time, as it sweeps farther out from our brightening star, it will eventually leave Earth behind. To make matters worse, recent calculations suggest that Earth is not in the middle of the habitable zone but rather on the zone's inner cusp, already teetering close to the edge of overheating [*see box on page 110*].

Consequently, within about half a billion years our sun will be bright enough to give Earth a feverish climate that will threaten the survival of complex multicellular life. By some 1.75 billion years from now, the steadily brightening star will make our world hot enough for the oceans to evaporate, exterminating any simple life lingering on the surface. In fact, Earth is well past its habitable prime, and the biosphere is fast approaching its denouement. All things considered, it seems reasonable to say our planet is at present only marginally habitable.

SEEKING A SUPERHABITABLE WORLD

IN 2012 I FIRST BEGAN THINKING about what worlds more suitable to life might look like while I was researching the possible habitability of massive moons orbiting gas-giant planets. In our solar system, the biggest moon is Jupiter's Ganymede, which has a mass only 2.5 percent that of Earth—too small to easily hang on to an Earth-like atmosphere. But I realized that there are plausible ways for moons approaching the mass of Earth to form in other planetary systems, potentially around giant planets within their stars' habitable zones, where such moons could have atmospheres similar to our own planet.

Such massive "exomoons" could be superhabitable because they offer a rich diversity of energy sources to a potential biosphere. Unlike life on Earth, which is powered primarily by the sun's light, the biosphere of a superhabitable exomoon might also draw energy from the reflected light and emitted heat of its nearby giant planet or even from the giant planet's gravitational field. As a moon orbits around a giant planet, tidal forces can cause its crust to flex back and forth, creating friction that heats the moon from within. This phenomenon of tidal heating is probably what creates the subsurface oceans thought to exist on Jupiter's Europa and Saturn's moon Enceladus. That said, this energetic diversity would be a double-edged sword for a massive exomoon because slight imbalances among the overlapping energy sources could easily tip a world into an uninhabitable state.

No exomoons, habitable or otherwise, have yet been detected with certainty, although some may sooner or later be revealed by archival data from observatories such as NASA's Kepler space telescope. For the time being, the existence and possible habitability of these objects remain quite speculative.

Superhabitable planets, on the other hand, may already exist within our catalog of confirmed and candidate exoplanets. The first exoplanets found in the mid-1990s were all gas giants similar in mass to Jupiter and orbiting far too close to their stars to harbor any life. Yet as planet-hunting techniques have improved over time, astronomers have begun finding progressively smaller planets in wider, more clement orbits. Most of the planets discovered over the past few years are so-called super-Earths, planets larger than Earth by up to 10 Earth masses, with radii between that of Earth and Neptune. These planets have proved to be extremely common around other stars, yet we have nothing like them orbiting the sun, making our own solar system appear to be a somewhat atypical outlier.

Many of the bigger, more massive super-Earths have radii suggestive of thick, puffy atmospheres, making them more likely to be "mini Neptunes" than super-sized versions of Earth. But some of the smaller ones, worlds perhaps up to twice the size of Earth, probably do have Earth-like compositions of iron and rock and could have abundant liquid water on their surfaces if they orbit within their stars' habitable zones. A number of the potentially rocky super-Earths, we now know, orbit stars called M dwarfs and K dwarfs, which are smaller, dimmer and much longer-lived than our sun. In part because of the extended lives of their diminutive stars, these super-sized Earths are currently the most compelling candidates for superhabitable worlds, as I have shown in modeling work with my collaborator John Armstrong, a physicist at Weber State University.

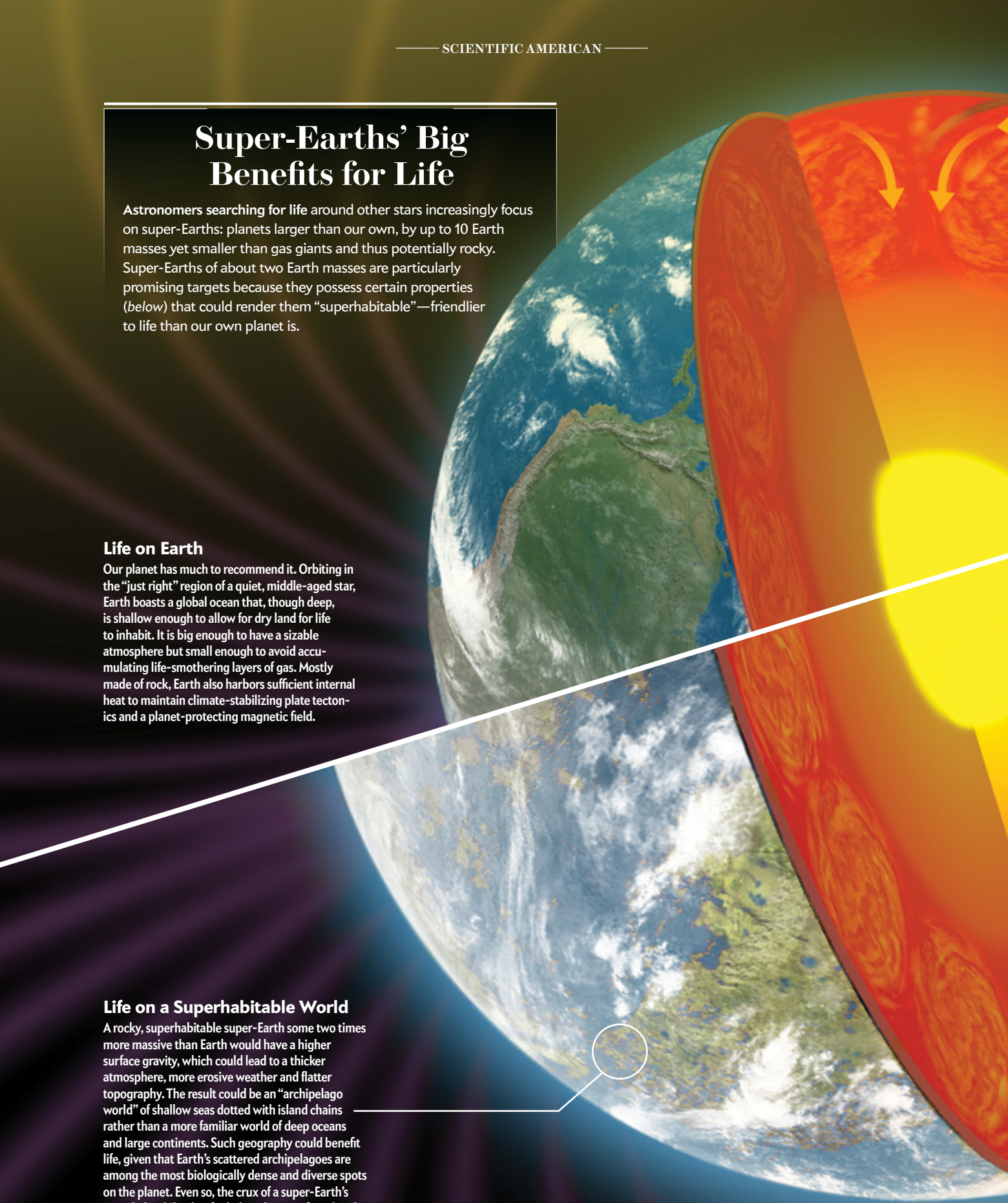
THE BENEFITS OF LONGEVITY

WE BEGAN OUR WORK with the understanding that a truly long-lived host star is the most fundamental ingredient for superhabitability; after all, a planetary biosphere is unlikely

© 2017 Scientific American

Super-Earths' Big Benefits for Life

Astronomers searching for life around other stars increasingly focus on super-Earths: planets larger than our own, by up to 10 Earth masses yet smaller than gas giants and thus potentially rocky. Super-Earths of about two Earth masses are particularly promising targets because they possess certain properties (*below*) that could render them "superhabitable"—friendlier to life than our own planet is.



Life on Earth

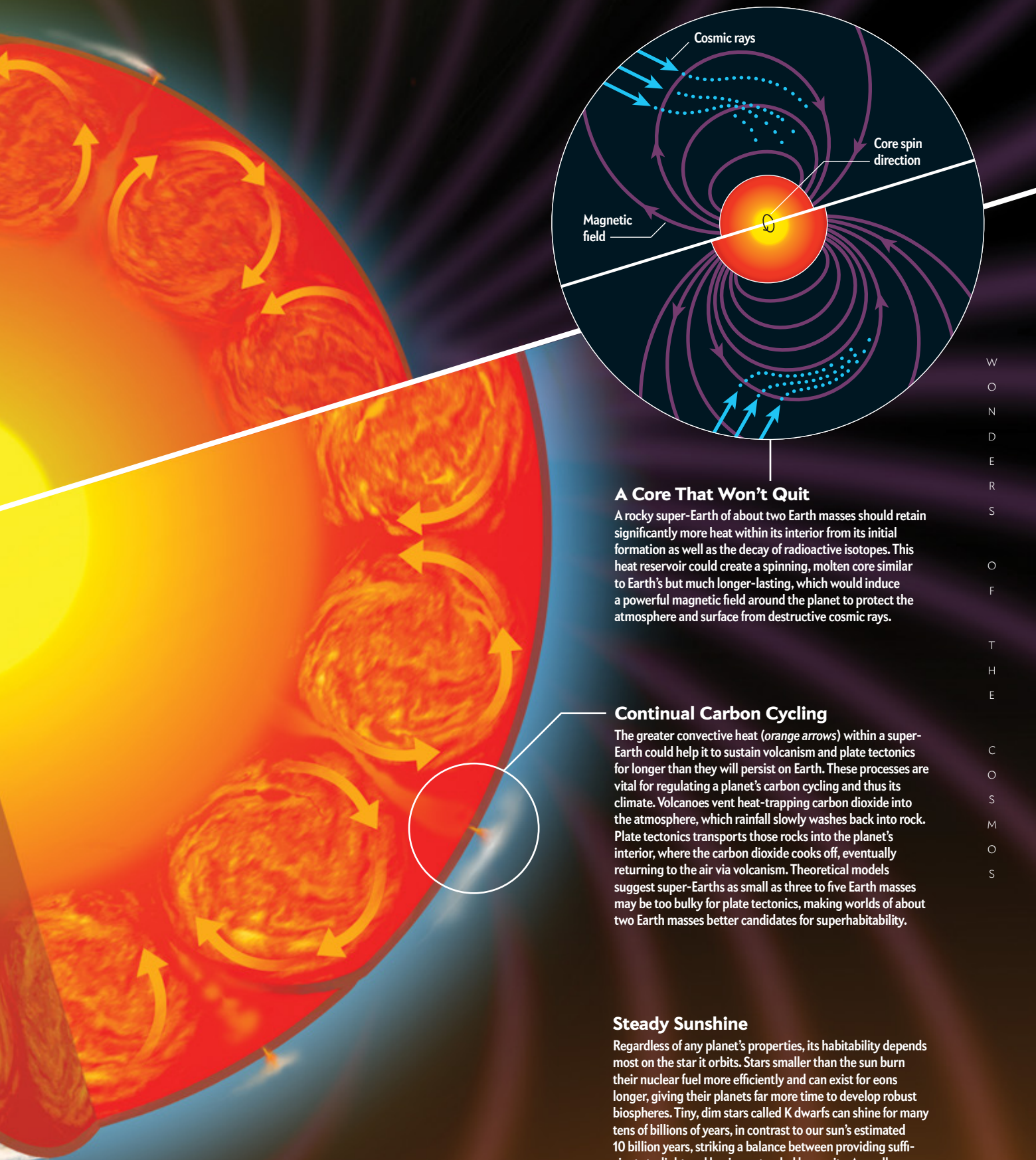
Our planet has much to recommend it. Orbiting in the "just right" region of a quiet, middle-aged star, Earth boasts a global ocean that, though deep, is shallow enough to allow for dry land for life to inhabit. It is big enough to have a sizable atmosphere but small enough to avoid accumulating life-smothering layers of gas. Mostly made of rock, Earth also harbors sufficient internal heat to maintain climate-stabilizing plate tectonics and a planet-protecting magnetic field.

Life on a Superhabitable World

A rocky, superhabitable super-Earth some two times more massive than Earth would have a higher surface gravity, which could lead to a thicker atmosphere, more erosive weather and flatter topography. The result could be an "archipelago world" of shallow seas dotted with island chains rather than a more familiar world of deep oceans and large continents. Such geography could benefit life, given that Earth's scattered archipelagoes are among the most biologically dense and diverse spots on the planet. Even so, the crux of a super-Earth's superhabitability lies far below these surface details, in the planet's interior (*opposite page*).

Illustration by Ron Miller (planets) and Jen Christiansen (inset)

© 2017 Scientific American



A Core That Won't Quit

A rocky super-Earth of about two Earth masses should retain significantly more heat within its interior from its initial formation as well as the decay of radioactive isotopes. This heat reservoir could create a spinning, molten core similar to Earth's but much longer-lasting, which would induce a powerful magnetic field around the planet to protect the atmosphere and surface from destructive cosmic rays.

Continual Carbon Cycling

The greater convective heat (*orange arrows*) within a super-Earth could help it to sustain volcanism and plate tectonics for longer than they will persist on Earth. These processes are vital for regulating a planet's carbon cycling and thus its climate. Volcanoes vent heat-trapping carbon dioxide into the atmosphere, which rainfall slowly washes back into rock. Plate tectonics transports those rocks into the planet's interior, where the carbon dioxide cooks off, eventually returning to the air via volcanism. Theoretical models suggest super-Earths as small as three to five Earth masses may be too bulky for plate tectonics, making worlds of about two Earth masses better candidates for superhabitability.

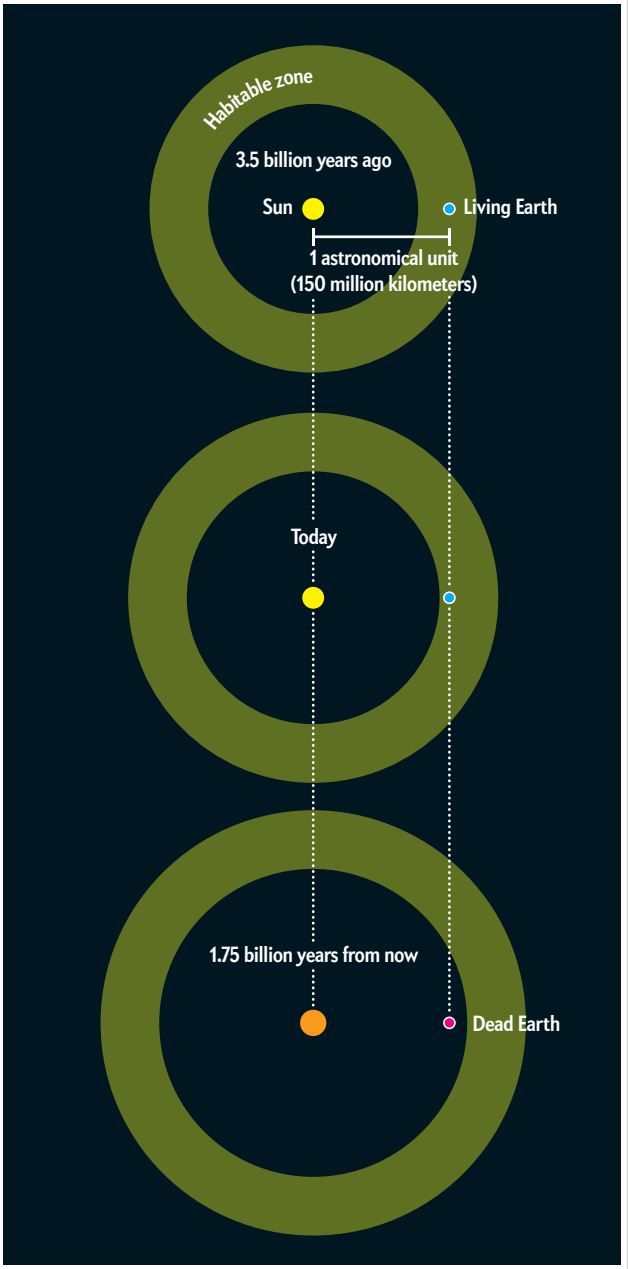
Steady Sunshine

Regardless of any planet's properties, its habitability depends most on the star it orbits. Stars smaller than the sun burn their nuclear fuel more efficiently and can exist for eons longer, giving their planets far more time to develop robust biospheres. Tiny, dim stars called K dwarfs can shine for many tens of billions of years, in contrast to our sun's estimated 10 billion years, striking a balance between providing sufficient starlight and having extended longevity. A small super-Earth orbiting in the habitable zone of a K dwarf may reside in the sweet spot of superhabitability.

© 2017 Scientific American

Aging Stars Turn Up the Heat on Habitable Planets

On human timescales, a star's habitable zone appears to be static. But because stars brighten as they age, over eons the zone sweeps outward, eventually leaving living worlds behind. Earth is poised near the inner edge of the sun's habitable zone and will become too hot to harbor liquid water in some 1.75 billion years. Smaller stars shine dimmer and longer than the sun, scarcely budging their habitable zones over tens of billions of years, potentially extending their planets' lives.



to survive its sun's demise. Our sun is 4.6 billion years old, approximately halfway through its estimated 10-billion-year lifetime. If it were slightly smaller, however, it would be a much longer-lived K dwarf star. K dwarfs have less total nuclear fuel to burn than more massive stars, but they use their fuel more efficiently, increasing their longevity. The middle-aged K dwarfs we observe today are billions of years older than the sun and will still be shining billions of years after our star has expired. Any potential biospheres on their planets would have much more time in which to evolve and diversify.

A K dwarf's light would appear somewhat ruddier than the sun's, as it would be shifted more toward the infrared, but its spectral range could nonetheless support photosynthesis on a planet's surface. M dwarf stars are smaller and more parsimonious still and can steadily shine for hundreds of billions of years, but they shine so dimly that their habitable zones are very close-in, potentially subjecting planets there to powerful stellar flares and other dangerous effects. Being longer-lived than our sun yet not treacherously dim, K dwarfs appear to reside in the sweet spot of stellar superhabitability.

Today some of these long-living stars may harbor potentially rocky super-Earths that are already several billion years older than our own solar system. Life could have had its genesis in these planetary systems long before our sun was born, flourishing and evolving for billions of years before even the first biomolecule emerged from the primordial soup on the young Earth. I am particularly fascinated by the possibility that a biosphere on these ancient worlds might be able to modify its global environment to further enhance habitability, as life on Earth has done. One prominent example is the Great Oxygenation Event of about 2.4 billion years ago, when substantial amounts of oxygen first began to accumulate in Earth's atmosphere. The oxygen probably came from oceanic algae and eventually led to the evolution of more energy-intensive metabolisms, allowing creatures to have bigger, more durable and active bodies. This advancement was a crucial step toward life's gradual emergence from Earth's oceans to colonize the continents. If alien biospheres exhibit similar trends toward environmental enhancement, we might expect planets around long-lived stars to become somewhat more habitable as they age.

To be superhabitable, exoplanets around small, long-lived stars would need to be more massive than Earth. That extra bulk would forestall two disasters most likely to befall rocky planets as they age. If our own Earth were located in the habitable zone of a small K dwarf, the planet's interior would have grown cold long before the star expired, inhibiting habitability. For example, a planet's internal heat drives volcanic eruptions and plate tectonics, processes that replenish and recycle atmospheric levels of the greenhouse gas carbon dioxide. Without those processes, a planet's atmospheric CO_2 would steadily decrease as rainfall washed the gas out of the air and into rocks. Ultimately the CO_2-dependent global greenhouse effect would grind to a halt, increasing the likelihood that an Earth-like planet would enter an uninhabitable "snowball" state in which all of its surface water freezes.

Illustration by Jen Christiansen

© 2017 Scientific American

Beyond the potential breakdown of a planet-warming greenhouse effect, the cooling interior of an aging rocky world could also cause the collapse of any protective planetary magnetic field. Earth is shielded by a magnetic field generated by a spinning, convecting core of molten iron, which acts like a dynamo. The core remains liquefied because of leftover heat from the planet's formation, as well as from the decay of radioactive isotopes. Once a rocky planet's internal heat reservoir became exhausted, its core would solidify, the dynamo would cease, and the magnetic shield would fall, allowing cosmic radiation and stellar flares to erode the upper atmosphere and impinge on the surface. Consequently, old Earth-like planets would be expected to lose substantial portions of their atmospheres to space, and higher levels of damaging radiation could harm surface life.

Rocky super-Earths as much as twice our planet's size should age more gracefully than Earth, retaining their inner heat for much longer because of their significantly greater bulks. But planets larger than about three to five Earth masses may actually be too bulky for plate tectonics because the pressures and viscosities in their mantles become so high that they inhibit the required outward flow of heat. A rocky planet only two times the mass of Earth should still possess plate tectonics and could sustain its geologic cycles and magnetic field for several billion years longer than Earth could. Such a planet would also be about 25 percent larger in diameter than Earth, giving any organisms about 56 percent more surface area than our world on which to live.

SUPERHABITABLE WORLDS ARE SLIGHTLY LARGER THAN EARTH AND ORBIT STARS THAT ARE SOMEWHAT SMALLER AND DIMMER THAN THE SUN.

LIFE ON A SUPERHABITABLE SUPER-EARTH

WHAT WOULD a superhabitable planet look like? Higher surface gravity would tend to give a middling super-Earth planet a slightly more substantial atmosphere than Earth's, and its mountains would erode at a faster rate. In other words, such a planet would have relatively thicker air and a flatter surface. If oceans were present, the flattened planetary landscape could cause the water to pool in large numbers of shallow seas dotted with island chains rather than in great abyssal basins broken up by a few very large continents [*see box on pages 108 and 109*]. Just as biodiversity in Earth's oceans is richest in shallow waters near coastlines, such an "archipelago world" might be enormously advantageous to life. Evolution might also proceed more quickly in isolated island ecosystems, potentially boosting biodiversity.

Of course, lacking large continents, an archipelago world would potentially offer less total area than a continental world for land-based life, which might reduce overall habitability. But not necessarily, especially given that a continent's central regions could easily become a barren desert as a result of being far from temperate, humid ocean air. Furthermore, a planet's habitable surface area can be dramatically influenced by the orientation of its spin axis with respect to its orbital plane around its star. Earth, as an example, has a spin-orbit axial tilt of about 23.4 degrees, giving rise to the seasons and smoothing out what would otherwise be extreme temperature differences between the warmer equatorial and colder polar regions. Compared with Earth, an archipelago world with a favorable spin-orbit alignment could have a warm equator as well as warm, ice-free poles and, by virtue of its larger size and larger surface area on its globe, would potentially boast even more life-suitable land than if it had large continents.

Taken together, all these thoughts about the features important to habitability suggest that superhabitable worlds are slightly larger than Earth and have host stars somewhat smaller and dimmer than the sun. If correct, this conclusion is tremendously exciting for astronomers because across interstellar distances super-Earths orbiting small stars are much easier to detect and study than twins of our own Earth-sun system. So far statistics from exoplanet surveys suggest that super-Earths around small stars are substantially more abundant throughout our galaxy than Earth-sun analogues. Astronomers seem to have many more tantalizing places to hunt for life than previously believed.

One of Kepler's prize finds, the planet Kepler-186f, comes to mind. Announced in April 2014, this world is 11 percent larger in diameter than Earth and probably rocky, orbiting in the habitable zone of its M dwarf star. It is probably several billion years old, perhaps even older than Earth. It is about 500 light-years away, placing it beyond the reach of current and near-future observations that could better constrain predictions of its habitability—but for all we know, Kepler-186f could be a superhabitable archipelago world.

Closer superhabitable candidates orbiting nearby small stars could soon be discovered by various projects, most notably the European Space Agency's PLATO mission, slated to launch by 2026. Such nearby systems could become prime targets for the James Webb Space Telescope, an observatory scheduled to launch in 2018, which will seek signs of life within the atmospheres of a small number of potentially superhabitable worlds. With considerable luck, we may all soon be able to point to a place in the sky where a more perfect world exists. SA

René Heller is a postdoctoral fellow at the Max Planck Institute for Solar System Research in Göttingen, Germany, and a member of the Canadian Astrobiology Training Program. His research focuses on the formation, orbital evolution, detection and habitability of extrasolar moons. He is informally known as the best German rice pudding cook in the world.

MORE TO EXPLORE

Habitable Climates: The Influence of Obliquity. David S. Spiegel, Kristen Menou and Caleb A. Scharf in *Astrophysical Journal,* Vol. 691, No. 1, pages 596–610; January 20, 2009. http://iopscience.iop.org/0004-637X/691/1/596/article

Exomoon Habitability Constrained by Illumination and Tidal Heating. René Heller and Rory Barnes in *Astrobiology,* Vol. 13, No. 1, pages 18–46; 2013. http://arxiv.org/abs/1209.5323

Habitable Zone Lifetimes of Exoplanets around Main Sequence Stars. Andrew J. Rushby, Mark W. Claire, Hugh Osborn and Andrew J. Watson in *Astrobiology,* Vol. 13, No. 9, pages 833–849; September 18, 2013.

Superhabitable Worlds. René Heller and John Armstrong in *Astrobiology,* Vol. 14, No. 1, pages 50–66; January 16, 2014. http://arxiv.org/abs/1401.2392

scientificamerican.com/magazine/sa

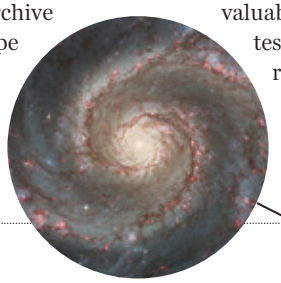
© 2017 Scientific American

END NOTE

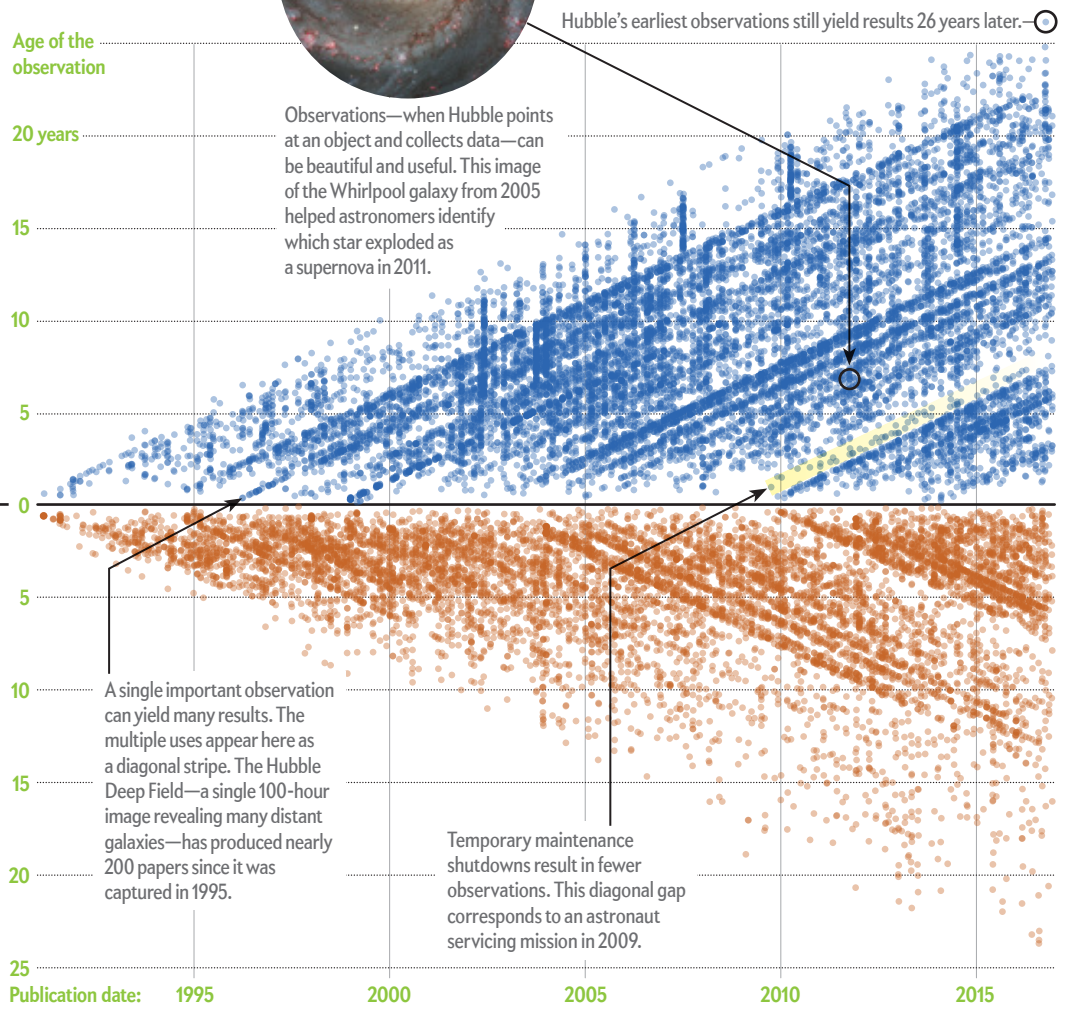
Long Live Hubble

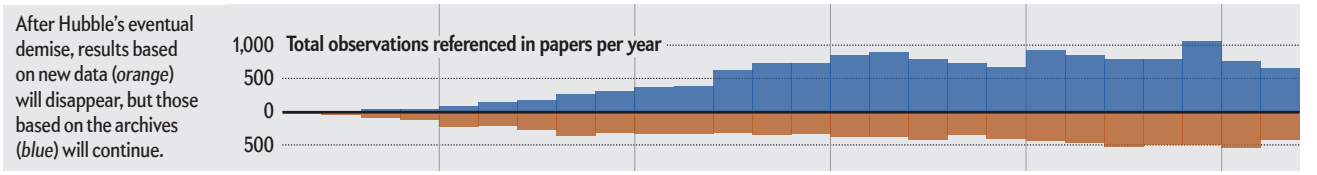
Data from the space telescope will yield discoveries long after the instrument is kaput

With luck, the Hubble Space Telescope will yield more startling images of the cosmos for years. But because NASA is no longer servicing the telescope, it is expected to give up the ghost sometime after 2020. That does not mean that the Hubble discoveries will stop, however. NASA maintains an archive of data gathered over Hubble's lifetime—the telescope went live in 1991—and makes it available to the public for free. The archive has already yielded discoveries such as nebulae and distant galaxies. "The legacy is a treasure trove of data that can be mined in the future," says Arfon Smith, who leads a new data-science initiative at Hubble's home, the Space Telescope Science Institute. "The data are incredibly valuable and still ridiculously useful." The archive is a testament to the enduring value of big, basic science research—the data can pay off in ways astronomers cannot yet imagine. *—Katie Peek*



Observations—when Hubble points at an object and collects data—can be beautiful and useful. This image of the Whirlpool galaxy from 2005 helped astronomers identify which star exploded as a supernova in 2011.


How to Read It
Each dot represents a Hubble observation referenced in a published paper.
Blue dots (nearly 14,000) relate to Hubble's archive: instead of asking for new observations, the researcher answered a question using existing images.
Orange dots (nearly 9,000) relate to new data: a researcher makes a specific observation request to answer a question.
An additional 10,000 references to Hubble observations appeared in papers using a mix of the telescope's new and archival images. Data are current as of early December 2016.
Age of the observation
20 years
15
10
5
0
5
10
15
20
25
Hubble's earliest observations still yield results 26 years later.
A single important observation can yield many results. The multiple uses appear here as a diagonal stripe. The Hubble Deep Field—a single 100-hour image revealing many distant galaxies—has produced nearly 200 papers since it was captured in 1995.
Temporary maintenance shutdowns result in fewer observations. This diagonal gap corresponds to an astronaut servicing mission in 2009.
Publication date:
1995
2000
2005
2010
2015


After Hubble's eventual demise, results based on new data (orange) will disappear, but those based on the archives (blue) will continue.
Total observations referenced in papers per year
1,000
500
0
500

SOURCE: BARBARA A. MIKULSKI ARCHIVE FOR SPACE TELESCOPES, SPACE TELESCOPE SCIENCE INSTITUTE (*data*); HUBBLE HERITAGE TEAM (STScI/AURA) (*galaxy*)

Graphic by Katie Peek

© 2017 Scientific American

BRIGHT HORIZONS 34

SCIENTIFIC AMERICAN Travel

Panama Canal | January 20th – February 4th, 2018

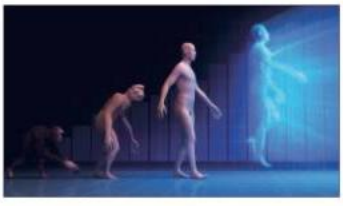
Sail from Atlantic to Pacific through an iconic human endeavor, transiting the Panama Canal on the Bright Horizons cruise conference. Voyage from San Diego to Ft. Lauderdale on Holland America's ms Westerdam. Visit regions of rain forest, colonial architecture, beautiful beaches, and coffee and cacao farms.

While at sea we'll discuss the latest in science and technology. Explore ports as you wish venturing into wildlife-rich rain forests, visiting the ancient home of chocolate, or relaxing on mellow beaches.

Cruise Central America with Scientific American and enrich your knowledge of important issues. Make your reservation today!

Cruise prices vary from $1,939 for an Interior Stateroom to $15,199 for a Pinnacle Suite, per person (pp). For those attending our SEMINARS, there is a $1,575 fee. Add'l pp fees: gov't taxes and fees ($410), non-refundable Booking Service Fee ($150), tour leader gratuities ($10 per day), and onboard gratuities (for your cabin steward and dining room staff, approx $11.50 per day). The Program, cruise pricing, and options are subject to change. For more information email us at info@InsightCruises.com.

SEMINARS



ANTHROPOLOGY:

- Human Evolution: the Big Picture
- The First Humans
- The Neanderthals: Another Kind of Human
- The Rise of Homo sapiens

THE UNIVERSE:

- Planets Everywhere
- Albatrosses, Beetles, and Cetaceans
- Jupiter and Saturn Revisited
- In Defense of Crazy Ideas

COSMOLOGY:

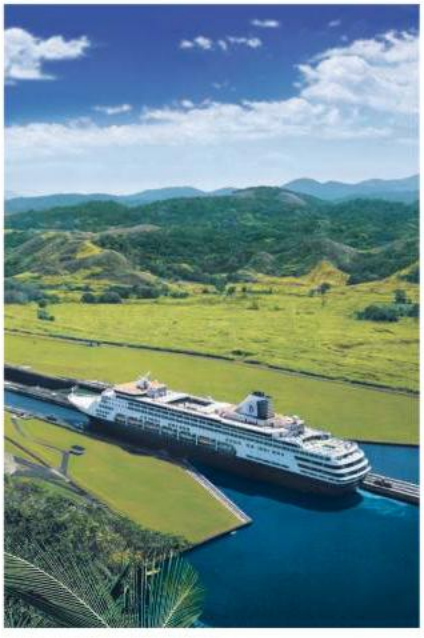
- The Origin of Our Universe
- A Brief History of Biological and Artificial Intelligence
- Mysteries of Our Universe
- The Future of Life, the Universe, and Everything

ARCHITECTURE, ENGINEERING, AND CONSTRUCTION:

- A Field Guide to Great Structures
- Strength through Curvature
- The Canal as an Engineered System
- Construction of the Panama Canal

SPEAKERS:

Stephen Ressler, Ph.D.
David Stevenson, Ph.D.
Chris Stringer, Ph.D.
Max Tegmark, Ph.D.



FOR MORE INFO Please email: **info@InsightCruises.com** or visit: **ScientificAmerican.com/travel**

CST# 2065380-40 °Scientific American, Inc.

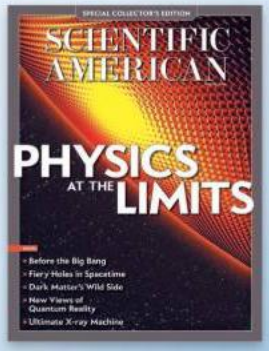
SCIENTIFIC AMERICAN™
Collections

Dive Deep into **Special Editions**

Explore over 50 single-topic special editions from *Scientific American* and *Scientific American MIND.*

From the Science of Dogs & Cats to Physics, our in-depth anthologies focus the lens on a distinct subject in fascinating detail. Previously available on newsstands, these special editions are now reissued in digital format on our website for you to explore.

Find the special edition for you at
scientificamerican.com/collections



Copyright © 2016 by Scientific American, a division of Nature America, Inc. All rights reserved.